

# Beyond HRV: Analysis of ECG Signals Using Attractor Reconstruction

Jane V Lyle<sup>1</sup>, Peter H Charlton<sup>2</sup>, Esther Bonet-Luz<sup>1</sup>, Gary Chaffey<sup>1</sup>, Mark Christie<sup>2</sup>, Manasi Nandi<sup>2</sup>, Philip J Aston<sup>1</sup>

<sup>1</sup> University of Surrey, Guildford, UK

<sup>2</sup> King's College London, London, UK

## Abstract

*Attractor reconstruction analysis has previously been applied to analyse arterial blood pressure and photoplethysmogram signals. This study extends this novel technique to ECG signals. We show that the method gives high accuracy in identifying gender from ECG signals, performing significantly better than the same classification by interval measures.*

## 1. Introduction

Traditional analysis of electrocardiogram (ECG) signals consists of identifying particular points and deriving lengths of various intervals from these points [1]. For instance, Heart Rate Variability (HRV) methods are used to analyse beat-to-beat intervals [2]. However, such approaches exclude the wealth of diagnostic information contained in the remainder of the waveform shape.

Our novel attractor reconstruction (AR) method uses all of the available data and so retains the underlying waveform information. It therefore goes “beyond HRV” and can characterise changes in signal morphology that are not detected by HRV [3]. AR analysis has previously been applied to arterial blood pressure (ABP) [3, 4] and photoplethysmogram (PPG) [5] signals, where it has been shown to supplement standard cardiovascular assessment.

The aim of this study is to demonstrate how the AR method can be applied to ECG signals. We illustrate this by comparing the performance of our AR method with standard interval analysis for identifying gender from ECG signals, both in a normal state and when dosed with drugs that impact ventricular repolarisation. It is well known that there are significant differences in various ECG parameters between females and males, including the PR interval, QRS duration and QT interval [6, 7]. Normal interval ranges may be gender-specific and certain cardiovascular diseases have a higher prevalence in one sex. Differences in ventricular repolarisation are of key interest, especially in the development and application of drugs that have a QT-prolonging effect [8].

## 2. Attractor Reconstruction Analysis

The attractor reconstruction method has been described previously [3–5], so only an overview is given here. The AR approach provides a means of analysing approximately periodic signals that may be irregular, strongly non-stationary and noisy, supporting its application to ECG signals. Moreover, it uses all of the data and so can detect changes in the morphology of the signal.

The first step of the AR method consists of embedding the one-dimensional ECG time series, which we denote by  $x(t)$ , in a three-dimensional phase space by using Takens' delay coordinate method [9]. From the original time series, we generate two further time series  $y(t) = x(t - \tau)$  and  $z(t) = x(t - 2\tau)$ , where  $\tau > 0$  is a fixed time delay, taken as one third of the average cycle length of the data, (i.e. one third of the cardiac cycle duration). A plot of  $(x, y, z)$  gives the reconstructed attractor in the bounded three-dimensional phase space (see Fig. 1(ii)).

Baseline variation, e.g. due to respiration or motion, in  $x(t)$  is removed by projecting the attractor onto a plane perpendicular to the vector  $(1, 1, 1)$  which we parametrise by the variables  $v = (x+y-2z)/\sqrt{6}$  and  $w = (x-y)/\sqrt{2}$ . Plotting  $(v, w)$  generates a two dimensional attractor as shown in Fig. 1(iii). Deriving a density from this attractor provides information about areas that are visited more or less frequently as shown in Fig. 1(iv).

Our method allows us to handle large data volumes. Individual features of attractors obtained from moving windows of data can be used to generate time traces which can indicate significant changes occurring in the signal.

The attractor of a normal lead II signal, as shown in Fig. 1(iv), has three long arms, each corresponding to the traverse of a given R peak by one of the points  $x$ ,  $y$  or  $z$ . The higher densities seen on these arms near the centre are a result of the T wave. Three shorter arms lying between the longer ones may be observed due to deeper Q or S waves.

Taking the attractor as a whole, we can derive measures related to its density distribution, overall size and symmetry. We can also consider the attractor centre and arms separately, as these relate to different aspects of the signal.

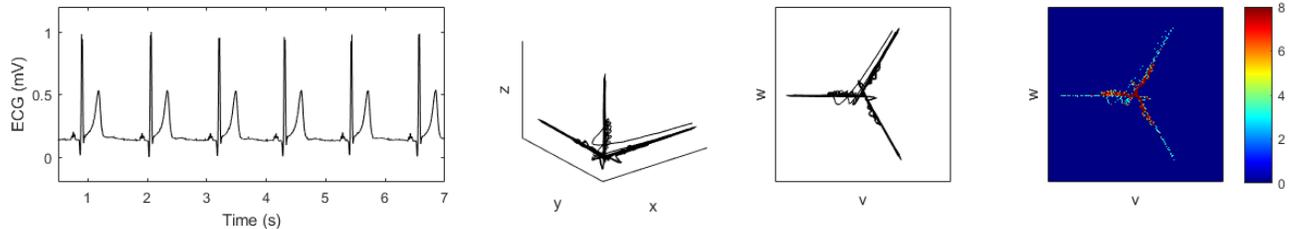


Figure 1. Attractor reconstruction (AR) performed on ECG signals. From left to right: (i) Lead II ECG signal, (ii) Three-dimensional attractor obtained using Takens’ delay coordinates, (iii) Two-dimensional attractor obtained by projecting the three-dimensional attractor onto a plane perpendicular to the vector (1,1,1), (iv) The two-dimensional attractor density.

### 3. Methods

This study demonstrates the application of the AR method to ECG signals, and exemplifies the attractor measures by gender classification of an appropriate dataset.

#### 3.1. Clinical dataset and data processing

Records for 22 individuals (aged 19 to 35 years, 11 female) were taken from the Physionet ecgdrvq database [10], providing a total of 5,232 12 lead ECG recordings. The original study is described in detail in [11].

Baseline ECG recordings were taken prior to a dose of one of five treatments (dofetilide, quinidine, ranolazine, verapamil or placebo), and 15 subsequent recordings were made over 24 hours. This was repeated weekly over 5 weeks for each treatment. Analysis in the original study included derivation of RR, PR, QRS,  $J-T_{\text{peakC}}$ ,  $T_{\text{peak}}-T_{\text{end}}$  and QTc intervals (taking  $J-T_{\text{peakC}}$  as  $J-T_{\text{peak}}/RR^{0.58}$  and QTc by Fridericia’s correction). The end of the T wave was calculated from the vectorcardiogram, generated from all the independent ECG leads [11].

The dataset provides 327 pre-treatment records. These were considered as baseline measurements and used to establish the AR method for healthy, unmedicated individuals. The remaining records were retained to demonstrate the application of AR measures to signals that may be undergoing significant changes due to drug administration.

The ECG signals were short and of good quality, so each 10 second recording was considered as a single window creating one attractor. Data was normalised to lie between 0 and 1. No additional filtering or data selection was applied in order to demonstrate that AR is a robust method.

The AR method was used to create the attractor and generate 56 measures for every lead of each record. Six measures characterising the attractor shape, density and symmetry were selected for the gender classification. As a comparison, we defined the six interval measures as RR, PR, QRS,  $J-T_{\text{peakC}}$ ,  $T_{\text{peak}}-T_{\text{end}}$  and QTc for each record [6, 7]. Since the interval measures had been obtained from all the independent leads of the ECG, we selected lead V2

for gender classification by machine learning as it could be expected to contain the most complete T wave information.

#### 3.2. Machine learning

A binary support vector machine (SVM) was selected to provide the classification. Binary SVM is a common technique that aims to maximise the distance between the support vectors of the two classes under consideration, thus being able to separate them into two distinct groups [12].

A cross-validation approach was applied. The complete records of one individual were removed from the dataset, and the remaining records were used to train an SVM classifier before testing it on the records of the “unknown” individual. This was then repeated for all individuals in turn, resulting in an average classification accuracy.

The first analysis was undertaken on baseline records only, both for training and testing. The second analysis used the initial baseline data for training but tested the full set (i.e. baseline and dosed) of the “unknown” individual’s records. Each analysis was undertaken for the AR measures and then the interval measures in order to compare their performance in classifying the records by gender.

### 4. Results

#### 4.1. Gender-specific measures

Attractors were reconstructed for each individual. Fig. 2(ii) shows typical female and male attractors. The arms are usually wider on female attractors, predominantly due to the width and morphology of the T wave, and reflecting a relatively shorter distance between the end of the T wave and the R peak in females. The upper arm (with  $w > 0$ ) usually shows the greatest variability in shape as a result of the labile shape of the T wave downstroke. In contrast, the relatively shorter QT interval on the male signal results in both the T and R peaks contributing to the same arm shape, thus creating an arm of increased uniformity and density.

Although these are distinct examples, the basis of these feature differences can be observed throughout the dataset,

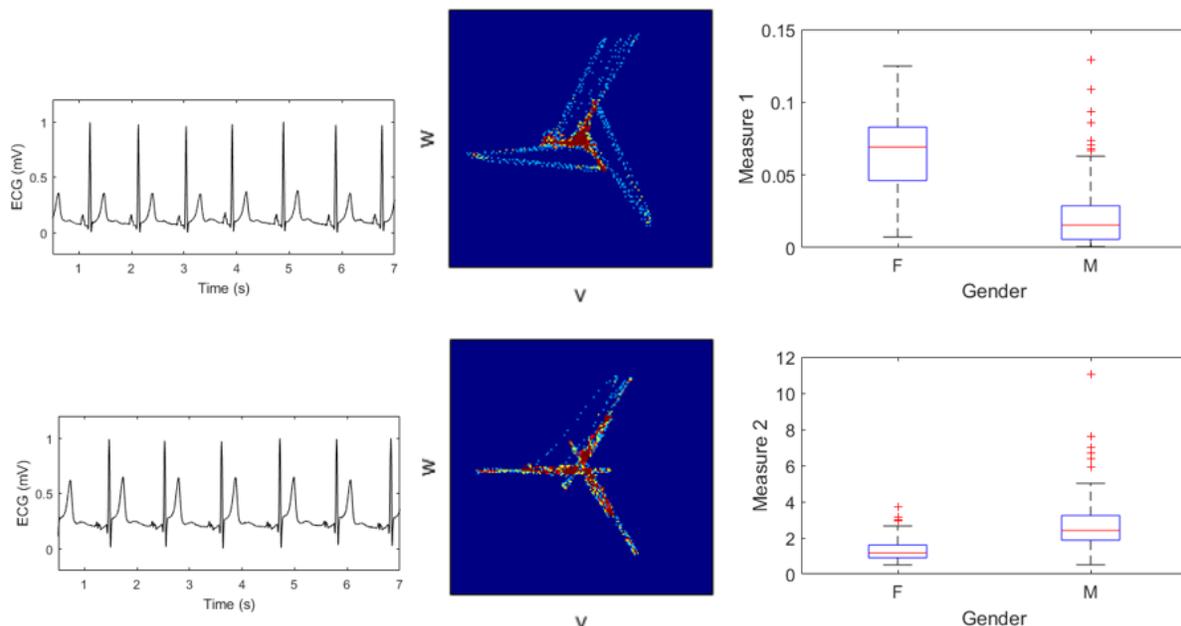


Figure 2. Gender differences in the attractor. From left to right: (i) Typical lead II ECG records for female (top) and male (bottom), (ii) The reconstructed attractor for each of these signals, (iii) Box plots for two different attractor measures for the 327 baseline records, showing a significant difference (two-sample t-test,  $p < 0.001$ ) between the genders.

and can be evaluated using the attractor measures. Fig. 2(iii) shows box plots for two of these measures. Each measure individually is significantly different between genders (two-sample t-test,  $p < 0.001$ ).

#### 4.2. Classification of unknown individuals

A support vector machine was used to classify the baseline records of “unknown” individuals by gender, using a cross-validation approach. Applying the AR measures achieved 96.3% accuracy, whereas the comparable interval measures correctly classified 77.7% of records. As Table 1 shows, the AR measures also had a more consistent classification, with a considerably lower standard deviation of accuracy across individuals. Conversely, the interval measures showed a high variability in accuracy, with some individuals’ records being classified with a high accuracy, whereas other individuals’ records were poorly identified.

#### 4.3. Classification of dosed individuals

The baseline records are only a small subset of the dataset. For the signals from the dosed individuals, the original study showed substantial T wave morphology changes for three of the drugs, and it might be expected that these changes could complicate gender identification.

However, using the SVM model trained only on the baseline data, the cross-validation classification of the complete (i.e. baseline and dosed) records of an “un-

Table 1. SVM classification for AR measures and interval measures, including the range and standard deviation (SD) of accuracy at an individual level. Sensitivity and specificity are with respect to female gender identification.

	AR measures (range, SD)	Interval measures (range, SD)
<b>Baseline (n=327)</b>		
Accuracy	96.3% (26.7, 7.3)	77.7% (86.7, 24.5)
Sensitivity	93.8%	80.9%
Specificity	98.8%	74.5%
<b>All (n=5,232)</b>		
Accuracy	93.1% (35.0, 8.7)	74.2% (86.3, 25.4)
Sensitivity	96.6%	77.2%
Specificity	89.5%	71.3%

known” individual using the AR measures still gave a high accuracy at 93.1%, as shown in Table 1. Interval measures classified with 74.2% accuracy. Again, similar observations could be made regarding the higher variability of interval measures at an individual level.

## 5. Discussion and Conclusions

In this study we have demonstrated the application of the AR method to ECG signals. We observed that the ECG

attractor has quite different shape and structure from that derived from ABP or PPG signals [3–5] which therefore required the extraction of different features.

The attractor features provided a strong classification of gender in previously unseen individuals, and performed significantly better than the interval measures which have previously been shown to discriminate gender [6, 7]. Cross-validation was applied to allow each of the 22 individuals' records to be classified as an "unknown" dataset, and this showed that the AR features had a much higher level of consistency of accurate classification. Only a basic SVM technique was used; full analysis of machine learning tools could improve the high accuracy even further.

The interval measures require accurate identification of points on the ECG signal, from which distances are derived and then averaged, and are thus susceptible to signal noise. The AR approach effectively averages waveform detail by generating the attractor from a window of data before extracting features making it more robust to noise. AR feature extraction is also simpler and requires less manual review than an automated generation of interval measures.

Training the SVM on only a small amount of data avoided slow processing. Furthermore, it sought to emphasise that the gender differences can be clearly separated from learning on only a small number of training records. Extrapolating this supports the utility of such measures in the quick and robust classification of larger datasets.

The results obtained by training on normal baseline and then testing on dosed individuals indicates that some attractor features are invariant with respect to the changes resulting from drugs impacting on ventricular repolarisation. Further investigation here could characterise which features reflect the changes in the T wave morphology as a result of a dosing treatment, which may be able to improve the risk assessment of such drugs.

Finally, we address the rationale for taking gender identification as our example for the AR method, since this information is readily available in every medical record. Our motivation is twofold. Firstly, it provides a simple classification problem for which the performance of the AR method and standard interval measures can be compared. But more importantly, this approach may help us to understand in more detail which features of the ECG signal are different for females and males. While many variations are already known [6, 7], examining which features are used by the machine learning to distinguish between the two classes may provide more detailed information on these differences and may facilitate more personalised health-care management. This will be the subject of future work.

## Acknowledgements

Esther Bonet-Luz, Gary Chaffey and Peter Charlton are grateful for EPSRC Impact Acceleration Account funding

and Jane Lyle is grateful for an EPSRC PhD studentship (EP/M508160/1).

## References

- [1] Clifford GD. ECG statistics, noise, artifacts, and missing data. In: Clifford GD, Azuaje F, McSharry PE, editors. *Advanced Methods for ECG Analysis*. London: Artech House; 2006:55–99.
- [2] Acharya UR, Joseph P, Kannathal N, Lim C, Suri J. Heart rate variability: A review. *Med. Bio. Eng. Comput.* 2006;44:1031–1051.
- [3] Aston PJ, Nandi M, Christie M, Huang Y. Comparison of attractor reconstruction and HRV methods for analysing blood pressure data. *Comp. Cardio.* 2014;41:437–440.
- [4] Aston PJ, Christie M, Huang Y, Nandi M. Beyond HRV: Attractor reconstruction using the entire cardiovascular waveform data for novel feature extraction. Submitted to *Phys. Meas.*
- [5] Charlton P, Camporota L, Smith J, Nandi M, Christie MI, Aston P et al. Measurement of cardiovascular state using attractor reconstruction analysis. In *Signal Processing Conf (EUSIPCO), 2015 Proc. of the 23rd Europe. IEEE 2015*.
- [6] Moss AJ. Gender differences in ECG parameters and their clinical implications. *Ann. Noninvasive Electrocardiol.* 2010;15:1–2.
- [7] Mandic S1, Fonda H, Dewey F, Le VV, Stein R, Wheeler M, Ashley EA, Myers J, Froelicher VF. Effect of gender on computerized electrocardiogram measurements in college athletes. *Phys Sportsmed.* 2010;38(2):156–64.
- [8] Mayuga KA, Thattassery E, Taneja T, Karha J et al. Circadian and gender effects on repolarization in healthy adults: A study using harmonic regression analysis. *Ann. Noninvasive Electrocardiol.* 2010;15:3–10.
- [9] Takens F. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Mathematics 1981;898:366–381*.
- [10] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circ.* 2000;101(23):e215–e220.
- [11] Johannesen L, Vicente J, Mason JW, Sanabria C, Waite-Labott K, Hong M et al. Differentiating Drug-Induced Multichannel Block on the Electrocardiogram: Randomized Study of Dofetilide, Quinidine, Ranolazine, and Verapamil. *Clin. Pharmacol. Ther.* 2014;96(5):549–558.
- [12] Li H, Feng X, Cao L, Zhang C, Tang C et al. Heartbeat classification using different classifiers with non-linear feature extraction. *Trans. Inst. Meas. Cont.* 2016;38(9):1033–1040.

Address for correspondence:

Jane Lyle

Department of Mathematics, University of Surrey, Guildford, Surrey GU2 7XH, UK

j.lyle@surrey.ac.uk