Detecting Predisposition to Torsade De Points Using a PCA-Based Method

A Khawaja[1], G Butrous[2], O Doessel[1]

[1]University of Karlsruhe (TH), Karlsruhe, Germany
[2]Pfizer Global Research and Development, Sandwich, United Kingdom

Abstract

Torsade de points (TDP) is a form of polymorphic ventricular tachycardia. It is associated with alternation of T wave and prolongation of the QT interval. The primary objective of this work is to find characteristics of the T waves before and after TDP using Principal Component Analysis (PCA). PCA was applied on T wave of 60 normal 24-hour tapes and 10 TDP 24-hour tapes from different studies recorded during 'Dofetilide' clinical trials (Pfizer, Inc.). All signals were first conditioned by eliminating baseline wander, detecting their significant points and extracting T waves of each channel into a data matrix. Afterwards, for every zero-centred data matrix, a covariance matrix and its corresponding eigenvalues and eigenvectors were calculated. Then, every beat is explained in terms of the eigenvectors delivering scores that characterise individual T wave. Results showed that Standard deviation (SD) of PCA scores for TDP patients before TDP syndrome are clearly higher than in case of healthy subjects.

1. Introduction

Torsade de Pointes (TDP) is a life-threatening arrhythmia closely linked to abnormal cardiac repolarization. The original name of TDP comes from Frensh language and means "twisting of the points", since QRS complexes wing up and down around the isoelectric axis periodically and in a chaotic fashion changing their morphology from beat to beat. Moreover, it is often followed by sudden cardiac death. The delay in phase III of the action potential, which is mediated by the HERG encoding the major repolarizing potassium current I_{kr}, is the underlying basis for the rhythm disturbance of TDP. In other words, TDP is characteristic of the congenital long QT syndrome caused by mutations in the HERG gene. The dysrrhythmia is allowed to emerge because the prolonged period of repolarization and the inhomogeneity of repolarization time among myocardial fibres. Furthermore, HERG appears to be the main molecular target for drugs which cause QT prolongation [1]. Cardiac safety is now a major issue in new drug development, because there is increasing awareness that many non-antiarrhythmic drugs can prolong the QT interval and provoke TDP [1]. Although the precise mechanism of torsade de pointes has not been established, recent in vivo studies [2], prefused wedge studies [3], and clinical observations made with monophasic AP recordings [4] have presented evidence in support of the hypothesis that an early afterdepolarization-induced, triggered response initiates TDP but that the arrhythmia is maintained by a re-entrant mechanism. They demonstrated an enhanced propensity of cardiac myocytes to generate early afterdepolarizations (EADs) in response to factors that prolong the action potential duration (APD). This proposed mechanism has been challenged by another one, which is based on the association between dispersion of repolarization (DOR) and TdP, suggesting involvement of reentrant excitation. Because the I_{ks} current density of the midmyocardial cells (M cells) is relatively weak, they are more sensitive to many APD -prolongation conditions than epicardial and endocardial cells and they can play an important role in arrhythmias which are dependent on delayed cardiac repolarization, such as LQTS. Despite relative normalization of the M-cell APD on subsequent beats, reentry persisted as the leading edge of the wavefront propagated into the recovering tail of the circuit. Such dynamic M-cell APD adaptation undoubtedly accounted for the rapidly changing trajectory of the reentrant circuit producing the characteristic polymorphic ECG morphology of TDP. The presence of uniform propagation on the epicardium may explain the appearance of a monomorphic waveform configuration in certain ECG leads but not others. Taken together, these findings suggest the existence of a single rotor during TDP that initially forms in the transmural wall and subsequently meanders into deeper layers of myocardium [5]. TDP is associated normally with marked prolongation of QT interval to 600 ms or greater. It is also associated with progressive changes in T-wave morphology and a bizarre shape of T wave especially after premature beats of long short cycle. Moreover, T wave alternans (TWA) is very important prog-

nostic indicator in that it is commonly observed just preceding episodes of TDP [6]. Inevitably, QTc interval prolongation has come to be recognized as a surrogate marker of the risk of TDP. Although it is the best and the simplest clinical measure that is available at present, QTc interval is not a reliable surrogate of TDP [7]. Furthermore, although the concept of QT dispersion is the best known and most widely investigated, it has also proved to be the least successful in predicting the risks of drug-induced TDP [7]. Monitor carefully the T wave morphology (TWM) changes in beat-to-beat manner appears to play a more important role in access the electrical stability of the ventricles and in detecting predisposition to TDP. That is, analysing the beat-to-beat changes and variability in TWM seems to be a robust precursor to TDP. In this paper, Principal Component Analysis (PCA) was applied on T waves of 60 normal and 10 TDP two-channel tapes from different studies recorded during Dofetilide clinical trials (Pfizer, Inc.). First, a new robust Discrete Wavelet Transformation-based (DWT-based) approach has been devised here to eliminate the artefacts of baseline wander and low-frequency components from the every channel of ECG signals [8]. The fiducial points, namely QRS complex onset, R peaks and T wave offset, for all beats in the same channel are then detected using an accurate ECG delineator [9]. After extracting QRST complexes (all beats from QRS complex onset till T-wave offset) from the ECG data set, outliers and premature beats are excluded from detected beats by means of Hotelling's T squared measure and they are not included in any further analysis. Afterwards, the outlier-free data were aligned very precisely to their R peaks by means of the cross-correlation coefficients and underwent a second-order Butterworth filter with cut-off freqeuncy of 20 Hz. Finally, T waves were considered only to build the input data matrix for the further PCA analysis. PCA is a multivariate statistical technique that allows for the identification of key variables, or combinations of variables, in a multidimensional data set that best explains the small differences between individual observations. In other words, PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each successive component accounts for as much of the remaining variability as possible. Our approach is aimed to employ PCA scores, namely the first PCA scores which are derived from PCA coefficients and represent the highest degree of deviations from the mean for every T wave, in a beat-to-beat analysis. In other words, PCA is employed to extract morphological features represented by PCA scores for every input T wave signal. The beat-to-beat fluctuation of the first PCA scores represents the deviation of T wave morphology to the mean T wave. The first PCA scores are then analysed in order to assess the degree of variation for all T waves compared to their mean T wave, i.e. the beat-to-beat T-wave morphology variation throughout the whole channel. The same procedure was applied on healthy and TDP signals.

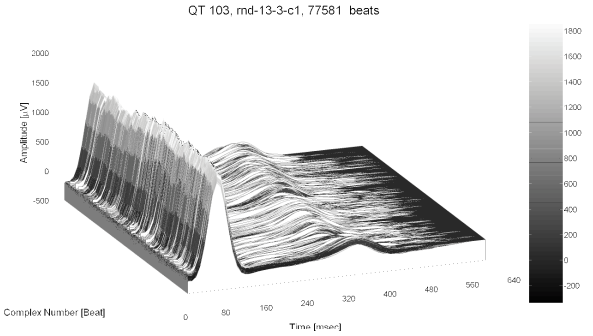

Figure 1. 77581 QRST complexes obtained by applying the data preconditioning on the first channel of a healthy tape (Pfizer, Inc.). The T-wave data matrix for PCA is the sub-matrix of QRST complexes starting at 200 (msec) in this case.

2. Methods

2.1. Data preconditioning

First of all, a new offline method for automatic baseline wander correction in Electrocardiogram, based on Mallat algorithm in DWT, is used to filter out the whole data set under studies. The method is able to segregate more than 99.5% of the baseline drift artefact in the signal without any distortion of ST segment as observed with other conventional high-pass filtering methods and other existing methods. Details of the method can be found in [8]. The next step is to localize QRS complex onset, R peak and T wave offset for every beat in the whole two-channel healthy and TDP tapes. An accurate and novel threshold-independent ECG delineation system was applied in this work [9]. Detection of the fiducial points is based completely on analysing the first scale details coefficients obtained by applying Haar function as mother wavelet. In order to get rid of unusual, ectopic and noisy detected beats and outliers, a new effective tool was developed to eliminate the wrong results. It provides a high level of confidence in the data to be analysed further. This method is based on Hotelling's T squared measure as an overall measure of variability in the dataset. After localizing QRST complexes, free of outliers or ectopic beats, in each channel, QRST complexes belonging to the same channel are extracted and assembled in one matrix, so that they represent the rows of that matrix . Afterwards, each QRST

complex in this matrix is shifted toward right and left a certain small number of samples and finally aligned at the position corresponding to the highest cross-correlation coefficient between this QRST complex and a chosen template signal, which is in this case the average of all QRST complexes. The aim of this fine alignment is to correct for any tiny misalignment between the extracted QRST complexes to their R peaks in order to provide a suitable input for the further analysis. Thereafter, a second-order Butterworth filter with cut-off freqeuncy of 20 Hz was applied on each QRST complex in the matrix. The input matrix for the next step, denoted as **T**, is a submatrix of QRST-complex matrix. It has the same number of rows (signals), but it has smaller number of columns as it starts from a chosen common point on ST segments from the QRST complexes and has the same end of the bigger matrix covering and including only T waves, figure 1.

2.2. Morphological feature extraction using PCA

1. *Organizing the data set*: Suppose that the matrix **T** is a training set with N samples and each sample T_i can be expressed by a row vector with the size of M as follows: $T_i = [T_{i1}, T_{i2}, \cdots, T_{iM}]$.
The training set is placed into a single matrix **T** of dimensions $N \times M$, so that N are the observations and M are the dimensions.
2. *Calculate the empirical mean* :The empirical mean along each dimension m = 1...M is calculated. Aftrewards, all computed mean values are placed into an empirical mean row vector u of dimensions M.
$u(m) = \frac{1}{N} \sum_{n=1}^{N} X(n,m), m = 1, 2, \cdots, M$
3. *Calculate the deviations from the mean*: The empirical mean row vector u is subtracted from each row of the data matrix T. Then a new mean-subtracted data matrix $B(N \times M)$ is derived.
$\mathbf{B} = \mathbf{T} - \mathbf{h} \cdot \mathbf{u}$,
where h is a column vector of ones and size of N x1 : $h(n) = 1 \qquad \text{for n} = 1 \ldots \text{N},$
4. *Find the covariance matrix*: As illustrated before, the $M \times M$ empirical covariance matrix **C** is calculated from the outer product of the zero-centered matrix B with itself: $\mathbf{C} = \mathbf{E}[\mathbf{B} \otimes \mathbf{B}] = \mathbf{E}[\mathbf{B} \cdot \mathbf{B}^*] = \frac{\mathbf{1}}{\mathbf{N-1}} \mathbf{B} \cdot \mathbf{B}^*$,
where E is the expected value operator,$\otimes$ is the outer product operator, and $*$ is the conjugate transpose operator.
5. *Find the eigenvectors and eigenvalues of the covariance matrix*: This step will typically require the use of a computer-based algorithm for computing the eigenvalue matrix **D** and the eigenvector matrix **V** of the covariance matrix **C**: $\mathbf{C} \cdot \mathbf{V} = \mathbf{V} \cdot \mathbf{D}$, Matrix **D** will take the form of an $M \times M$ diagonal matrix, where $\mathbf{D}[p,q] = \lambda_m \qquad \text{for} \qquad \text{p} = \text{q} = \text{m}$ is the m^{th} eigenvalue of the covariance matrix **C**, and $\mathbf{D}[p,q] = 0 \qquad \text{for} \qquad \text{p} \neq \text{q}.$
Matrix **V**, also of dimension $M \times M$, contains M column vectors, each of length M, which represent the M eigenvectors of the covariance matrix **C**.
The eigenvalues and eigenvectors are ordered and paired. The m_{th} eigenvalue corresponds to the m_{th} eigenvector.
6. *Rearrange the eigenvectors and eigenvalues*: The columns of the eigenvector matrix **V** and eigenvalue matrix **D** are sorted out in order of *decreasing* eigenvalues maintain the correct pairings between the columns in each matrix.
7. *Convert the source data to the new basis*: The new basis is denoted as *PCA-scores* or *the reconstruction parameter vectors* (RPV). The projected vectors are the columns of the matrix $\mathbf{Z}(\mathbf{N} \times \mathbf{M})$, namely $\mathbf{Z_{i1}}, \mathbf{Z_{i2}}, \cdots, \mathbf{Z_{iM}}$, where $i = 1 \ldots N$. The matrix **Z** is calculated by multiplying the eigenvector matrix with the zero-mean data matrix from the left as follows:

$$\mathbf{Z} = \mathbf{B} \cdot \mathbf{V} = \mathbf{KLT}\{\mathbf{T}\} = \begin{bmatrix} z_{11} \cdots z_{1M} \\ \vdots \ddots \vdots \\ z_{N1} \cdots z_{NM} \end{bmatrix},$$

The rows of **Z** correspond to the observations, whereas the columns refer to the components or dimensions.
In fact, the projected PCA-scores or vectors represent the Karhunen-Loève transform (KLT) of the data vectors in the columns of matrix **T**.

2.3. Analysing the first PCA scores

Since the first PCA scores represented in the first column vector of the matrix **Z** accounts the most of the variance in the data, analysis was carried out only on these scores so far. Standard deviation (SD) was used and applied on the first PCA scores as a simple linear measure to assess the beat-to-beat morphology variation. As the length of the first PCA score vector is N, the number of T waves in one channel, a window of length 60 with zero overlapping was chosen to scan this vector calculating the SD for the scores inside this window at each step. Finally, a series of SD values will come out. High SD value represents high beat-to-beat PCA score variation , i.e. high beat-to-beat T-wave morphology variation, and vice versa for the low SD value. In order to get an overall measure of variation for each channel, mean of the derived SD values was calculated.

3. Results

The method illustrated above for calculating the overall measure of variation was applied on every channel of all healthy subjects. Furthermore, this overall measure of variation was calculated twice on every channel of TDP

subjects. The first one measures T-wave overall variation from the beginning of the tape until TDP episode, whereas the second one measures T-wave overall variation starting after TDP episode until the end of the tape. The avarage value of the overall measure of variation for 84 useful channels from the **healthy tapes** was equal to **73.8119** $\pm$ **12.4222**. The avarage value of the overall measure of variation for T waves **before TDP episode** of 20 useful channels from the TDP tapes was equal to **241.9493** $\pm$ **168.3503**, whereas the avarage value of the overall measure of variation for T waves **after TDP episode** of 20 useful channels from the TDP tapes was equal to **145.7783** $\pm$ **86.3924**

Since the SD values for all TDP tapes before and after the episode are available, it is useful to employ them in order to have a closer and more detailed information about the beat-to-beat T wave morphology variation before and after TDP episode. Therefore, the mean of SD values for all channels before and after TDP episode was calculated. Because every channel has a different number of SD windows before and after TDP, the SD windows for all channels before TDP were rearranged, so that they are right aligned to the last window just before TDP episode. On the other hand, the SD windows for all channels after TDP were kept left aligned to the first window just after TDP episode. The calculation of the final mean SD values of similar windows did not take into calculation any possible missing SD values, i.e. missing windows . Figure 2 shows the final results.

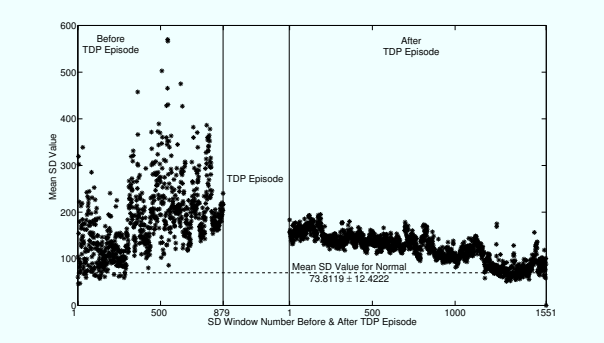

Figure 2. The final results obtained from 20 channels of 10 TDP tapes. TDP in all tapes occured and self-stopped, so that the patient got recovered again after its episode in all the cases under study.

4. Discussion and conclusions

Figure 2 shows that the beat-to-beat morphology T-wave variation before TDP episode is remarkably higher than the normal level, more chaotic and increased by approaching the TDP episode. Figure 2 illustrates also that the T-wave beat-to-beat morphology variation after TDP episode is not as high as before TDP episode and is decreasing by receding away from TDP episode until it reaches the normal variation level. Referring to our results, detecting predisposition to TDP is possible some hours prior to TDP episode employing the normal level of T-wave variation derived by our calculation.

References

[1] Tie HH. Cellular Mechanisms of QT prolongation and Proarrhythmia induced by non-antiarrhythmic drugs. University of New South Wales: Doctor in Medicine thesis, 2002.

[2] Verduyn SC, Vos MA, van der Zande J, Kulcsar A, Wellens HJ. Further observations to elucidate the role of interventricular dispersion of repolarization and early afterdepolarizations in the genesis of acquired torsade de pointes arrhythmias: a comparison between almokalant and d-sotalol using the dog as its own control. J Am Coll Cardiol Nov 1997; 30(6):1575–1584.

[3] Antzelevitch C, Sun ZQ, Zhang ZQ, Yan GX. Cellular and ionic mechanisms underlying erythromycin-induced long QT intervals and torsade de pointes. J Am Coll Cardiol Dec 1996;28(7):1836–1848.

[4] Shimizu W, Antzelevitch C. Cellular Basis for the ECG Features of the LQT1 Form of the Long-QT Syndrome : Effects of -Adrenergic Agonists and Antagonists and Sodium Channel Blockers on Transmural Dispersion of Repolarization and Torsade de Pointes. Circulation 1998;98(21):2314–2322.

[5] Akar FG, Yan GX, Antzelevitch C, Rosenbaum DS. Unique topographical distribution of M cells underlies reentrant mechanism of torsade de pointes in the long-QT syndrome. Circulation Mar 2002;105(10):1247–1253.

[6] Surawicz B, Fisch C. Cardiac alternans: diverse mechanisms and clinical manifestations. J Am Coll Cardiol 1992; 20(2):483–499.

[7] Shah RR. Drug-induced QT dispersion: does it predict the risk of torsade de pointes? J Electrocardiol Jan 2005; 38(1):10–18.

[8] Khawaja A, Sanial S, Dössel O. A wavelet-based technique for baseline wander correction in ecg and multi-channel ecg. In NBC05, 2005, IFMBE, volume 9. 2005; 291–292.

[9] Khawaja A, Sanial S, Dössel O. A wavelet-based multichannel ecg delineator. In EMBEC05, 2005, IFMBE, volume 11. 2005; .

Address for correspondence:

Antoun Khawaja
Universität Karlsruhe, Institute of Biomedical Engineering, Kaiserstrasse 12, Geb. 30.33, 76131 Karlsruhe, Germany.
Tel./Fax. : +49 (721) 608 -3851/-2789
Antoun.Khawaja@ibt.uni-karlsruhe.de