Waveform Detection in Holter ECG using Dynamic Time Warping

E. Vullings[1], J. García[2], and P. Laguna[2]

[1]Dept. Electrical Engineering, Delft Univ. of Technology, The Netherlands
[2]Dept. Ing. Electrónica y Comunicaciones, CPS, Univ. de Zaragoza, Spain

Abstract

We present Dynamic Time Warping as a real-time method to detect the onset and end of the different waves (P, QRS, and T) in the ECG during operations. The preliminary results based on an off-line evaluation using the QT database are compared to the annotations made by cardiologists and another method based on threshold detection. The current method gets comparable results for the mean error, although the standard deviation is larger than the other reported method.

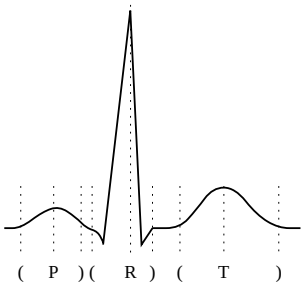


Figure 1: *Example of an ECG with its different fiducial points. The onset and offset of the different waveforms are annotated as '(' and ')', respectively.*

1 Introduction

The different wave intervals in ECG recordings (see Figure 1) contain some relevant clinical information [1]. This information should preferably be extracted in a fast and automatic way, so it can also be used as a source of information for an anesthetist during operations. There exist automatic, off-line methods based on thresholds that are used for the classification of wave onset and offset [1, 2], and here we will compare the results of the threshold based classifier, as well as the manual annotations of two independent cardiologists, with a classifier that uses Dynamic Time Warping (DTW). Instead of using predefined thresholds, this classifier takes the first labeled period as a reference, and matches all subsequent periods with this reference. Essentially, DTW aligns different periods in a nonlinear fashion, such that the total error will be minimal. After the alignment has been performed, the matching points (such as P, QRS, or T wave) will be labeled similarly. So, at the start of an operation, an anesthetist could be asked to label the important points of the ECG, and DTW will keep track of these points during the operation.

In any classification process, there are many types of error. In the case of labelling Holter ECG recordings, most problems are due to noise contaminating the signals, non-stationary or not well defined waveform morphologies, absence of certain waveforms, ambiguity of the waveform onset or offset, etc..

In order to evaluate the algorithm properly, use is made of the QT Database [3], which contains 105 excerpts of 15 minutes (sampled at 250 Hz), where 30 periods were labelled independently by two cardiologists. Every period contains the onset and offset of all waveforms that could be located by the cardiologists, as well as their fiducial points, if present.

2 Method

In order to label the fiducial points of the ECG, the first step is to extract all possible fiducial points. This step is performed by taking the Piecewise Linear Approximation (PLA) of the ECG (§2.1), which at the same time reduces the amount of data considerably. Instead of reinventing the wheel, a standard QRS detector is used for the detection of the QRS wave (§2.2). Next, the periods are extracted, and the first labeled PLA is used as a reference for subsequent processing (§2.3). Note that when the waveform morphology changes dramatically, the classification will be erroneous. However, this problem can be detected when using the distance measure obtained after the comparison with the DTW algorithm, as it will be much larger for different waveforms. Finally, the DTW algorithm is explained in more detail (§2.4), and the labels are placed (§2.5).

2.1 Piecewise Linear Approximation

To extract the fiducial points of the ECG, use is made of the Piecewise Linear Approximation proposed by Koski in [4]. It is fast, and only two parameters need to be set: the allowable perpendicular error ϵ, and

the step size s of the algorithm. The general idea is illustrated in Figure 2. Connect point i with point $i+s$. Compute the perpendicular error of the points with the line. If the maximum error is smaller than ϵ, connect i with $i+2s$, and continue. However, if the error is larger than ϵ, break the line at the point with maximum error j and compute the error for the line between i and j. Continue breaking the line until the maximum error is smaller than ϵ: the new end-point of the line will be j.

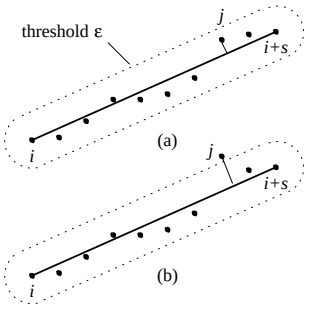


Figure 2: *Piecewise Linear Approximation of a set of points, where the dotted line represents the maximum acceptable error. In (a), this error is still acceptable, and more points can be added. In (b), however, the line has to be broken at point j.*

2.2 QRS Detection

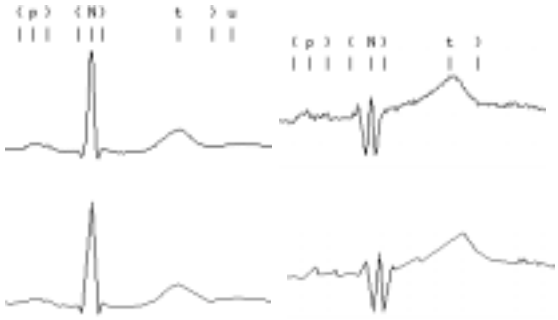
The QRS detection is based on ARISTOTLE [5]. Next, periods are created by dividing the period between two subsequent R peaks. Any PLA line that is at the division of these periods is cut in half as well.

2.3 Reference Period

With the PLA of every period available, the first annotated period in the QT database is selected as a reference for further comparisons. All annotated points are matched to the nearest location in the PLA. Note that it is not guaranteed that there will be an end point of a line near the annotation. However, as the PLA should have detected the fiducial points in the ECG, it is likely there is one (see Figure 3. If none is available, one could even decrease the maximum allowable error of the PLA and start again, until such a point is found, which is not done here.

2.4 Dynamic Time Warping

The principal part of the ECG segmentation is performed by Dynamic Time Warping (DTW), which aligns two different periods in a nonlinear way. Essentially, DTW is a child of a more general optimization method called dynamic programming. Its main



(a) Record sel103 (b) Record sel114

Figure 3: *Example of two ECG periods (upper figures), together with the automatically created annotations based on the PLA (lower figures).*

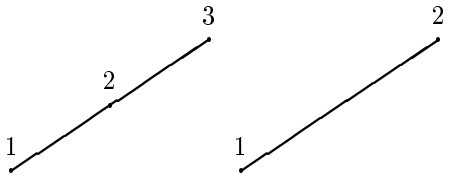
strength is in the comparison of different time sequences, either for alignment, or for calculating the distance (also called dissimilarity) between two sequences. A major advantage over other methods (such as the Euclidean distance) is the fact that the sequences do not need to have the same length. Sequences of different length are compressed or expanded until they match. Generally, the absolute values of the samples are used in DTW [6, 7]. Here, however, use is made of the slope values, which can be considered as a first-order DTW when compared to the zero-order DTW based on the absolute values [8]. The main difference between these two approaches can be best understood when studying Figure 4. Standard DTW will generate a dissimilarity when comparing Figure 4a and 4b, as the second point in (a) will have to be matched to one on the endpoints in (b). First-order DTW, however, will compare the two lines of (a), which have the same slope, with the single line in (b), without generating a dissimilarity. This is an advantage for the current problem, as it can often happen that a similar line segment is broken into two or more lines when generating a PLA, and first-order DTW allows it to be remapped to a single line without a large increase in the dissimilarity.

Given two sequences, the DTW algorithm, whether it is based on the absolute value or the slope, computes the optimal (in a certain sense) alignment in a recursive way. As an example, see Figure 5, where two short sequences are compared.

This illustrative example can be expressed as:

$$\mathbf{d}_{i,j} = \min \begin{cases} \mathbf{d}_{i-1,j} & + & \frac{1}{2}w(i,j)t(a_i) \\ \mathbf{d}_{i-1,j-1} & + & \frac{1}{2}w(i,j)(t(a_i + t(b_j)) \\ \mathbf{d}_{i,j-1} & + & \frac{1}{2}w(i,j)t(b_j) \end{cases} \quad (1)$$

Here, $d_{i,j}$ is the distance between the subsequences until points i and j, $w(i,j)$ is the dissimilarity between the two points or slopes, and $t(a_i)$ is the du-



(a) Two lines (b) Single line

Figure 4: *The difference between standard DTW (based on sample values) and first-order DTW (based on slopes) is illustrated here: the first will generate a dissimilarity when comparing (a) to (b), as the middle point has to be matched to an endpoint. First-order DTW, however, will match the slopes without generating a dissimilarity.*

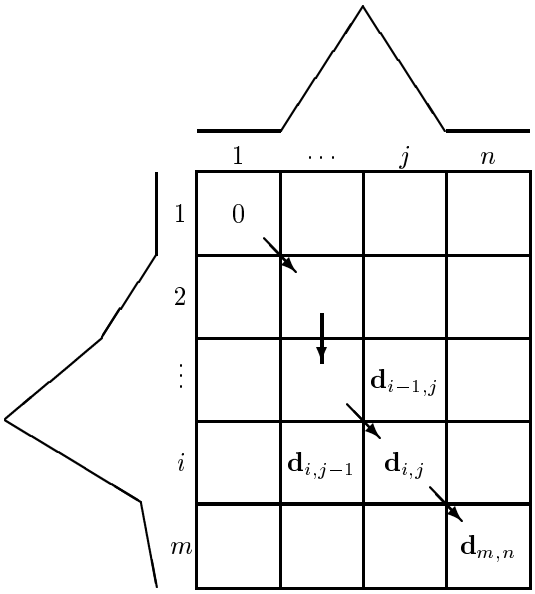


Figure 5: *Two short sequences are aligned to each other, starting in the upper-left corner, and working its way through to the lower-right corner. The optimal path is shown by arrows.*

ration of line i in sequence a. In this example, the durations of all lines are the same, but for the PLA this is naturally not the case. Of course, instead of using the PLA, one could also use the ECG directly, but this would increase the number of comparisons considerably. Another implemented reduction in the number of calculations is the possibility to compute the two parts of a sequence, one part before the QRS wave and one part after the QRS wave, independently. In that way, we only have to compute half of the original matrix.

As we are not so much interested in the final distance between two sequences, but want to know the alignment, we also have to construct a backtracking matrix $\mathbf{R}$. This matrix consist of pointers which show the path followed from cell $(1,1)$ to cell (m,n), where every entry in a cell corresponds to the minimum solution of (1). For the example given above, $\mathbf{R}$ would look like as is shown in Figure 6, where the pointers are replaced by arrows to make it more visual.

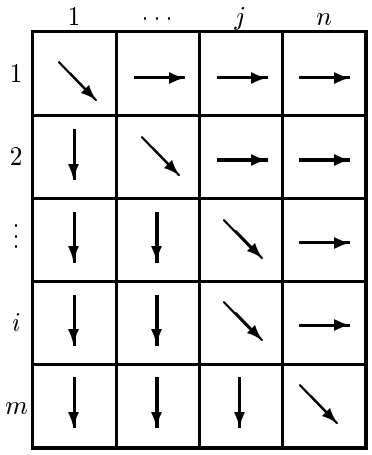


Figure 6: *The corresponding* $\mathbf{R}$ *matrix for the sequences of Figure 5. Starting from cell* (m,n) *backwards until cell* $(1,1)$ *gives the optimum alignment.*

2.5 Labeling the ECG

The final step, necessary for producing the output, is easy. Using the backtracking matrix $\mathbf{R}$, trace the matrix back starting from the final segment (m,n) to the first $(1,1)$. This gives the minimal path, or the optimal alignment between the two periods. Next, look for those lines in the reference period which were indicators of a fiducial point, and take the corresponding line of the new period. In the example shown above, we would label the top of the peak as cell $(3,2)$, so the time stamp applied would be after the second line (when we take the horizontal sequence as the reference). This time stamp together with the corresponding label of the fiducial point, is written to file.

3 Results

The obtained results are strongly influenced by the settings of the PLA parameters. A large accepted perpendicular error results in few points, which makes the precision of the onset and offset detection low. Therefore, we optimized these settings for different values of the parameters on five randomly chosen ECG records (which were excluded from further testing), The best results were obtained for $\epsilon = 2.5\mu V$ and $s = 48ms$. Note that the latter parameter is not very sensitive. A rule of thumb is twice the mean line length, which is also related to the compression ratio of the PLA.

Finally, a suitable function for $w(i,j)$ has to be chosen. Again, optimization between several options

for this function lead to $w(i,j) = (\arctan(s_A(i)) - \arctan(s_B(j)))^2$, where the $s_A(i)$ is the slope of line i of the first sequence. Likewise for $s_B(j)$. The use of the arctan makes the difference between steep lines smaller, so the QRS wave does not influence the total comparison too much.

With these settings we analyzed the QT dataset, excluding the files used for optimization, and the results are shown in Table 1. The results of the threshold segmentation method used in [2] are also shown here.

	DTW			Threshold			Car.
	beats	mean	SD	beats	mean	SD	SD
P_{on}	2345	-7.53	38.93	2596	10.26	14.08	10.2
P	2356	-5.43	32.02	2626	-0.48	10.96	
P_{end}	2364	-5.85	36.45	2627	-5.73	13.57	12.7
QRS_{on}	2612	-2.79	29.02	3130	-7.82	10.86	6.5
R	2613	0.28	13.05	3130	-9.32	4.41	
QRS_{end}	2613	-4.08	30.79	3130	-3.64	10.74	11.6
T_{on}	1065	-14.10	56.83	1241	-16.00	29.82	
T	2563	-5.75	58.36	2932	23.26	28.26	
T_{end}	2510	-0.90	65.72	2996	18.68	29.79	30.6

Table 1: *The first three columns are produced with DTW, the second three columns are reproduced from [2], and the last column shows the standard deviation for the cardiologists. Mean and standard deviation are in milliseconds.*

4 Conclusions and Discussion

From Table 1 it can be concluded that although the mean error is smaller when compared to the other automatic, threshold based annotator, the standard deviations are much larger. Therefore, improvements of the current method are necessary before it can be implemented. Further, tests should be performed to see whether the method is also suitable during an operation, where the signal quality can change very fast.

The fact that the standard deviation is much larger is probably due to false alignments occurring when there is much noise present in the signal. Also, the annotator in [2] used both signals, selecting the one with the best results compared to the cardiologist. Here, however, only the first signal was evaluated, although the cardiologists may have used the second signal for their annotations, if its signal-to-noise ratio (SNR) was better. Another possibility is that the PLA does not select the fiducial points properly, so the annotated points switch to the beginning and end of PLA line.

Several possibilities exist to improve the method: as in [1], we can also first try to filter the signal to improve the SNR, although the PLA already acts, more or less, as a low-pass filter. A decrease of the threshold around the fiducial points, so there is more detail around the important areas, could also make the detection more accurate. Further, the DTW can be constrained, so not all possible paths are evaluated, but only the more promising ones. Another improvement can be to compare the annotations provided by the cardiologist with the PLA of both signals, selecting that signal which fits the annotations best. Finally, one could switch between the two signals depending on their dissimilarity with the reference period, which is a byproduct of the DTW anyway. So, if the dissimilarity between the current period and the reference is smaller in signal 1, annotate this signal. Else, select the other.

References

[1] P. Laguna, R. Jané, and P. Caminal, "Automatic detection of wave boundaries in multilead ECG signals: validation with the CSE database", *Computers and biomedical research*, vol. 27, pp. 45—60, 1994.

[2] R. Jané, A. Blasi, J. García, and P. Laguna, "Evaluation of an automatic threshold based detector of waveform limits in holter ECG with the QT database", in *Computers in Cardiology*, 1997.

[3] P. Laguna, R. Mark, A. Goldberger, and G.B. Moody, "A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ecg.", in *Computers in Cardiology*, 1997.

[4] A. Koski, "Segmentation of digital signals based on estimated compression ratio", *IEEE trans. on Biomedical Eng.*, vol. 43, no. 9, pp. 928–938, 1996.

[5] G.B. Moody and R.G. Mark, "Development and evaluation of a 2-lead ECG analysis program", in *Computers in Cardiolog.* 1982, pp. 39—44, IEEE Computer Society Press.

[6] D.J. Berndt and J. Clifford, *Advances in Knowledge Discovery and Data Mining*, chapter 9, Finding patterns in time series: a dynamic programming approach, pp. 229–248, AAAI Press/The MIT Press, 1996.

[7] D. Sankoff and J.B. Kruskal, *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*, chapter 1 & 4, Addison-Wesley, 1983.

[8] H.J.L.M. Vullings, M.H.G. Verhaegen, and H.B. Verbruggen, "ECG segmentation using time-warping", in *Advances in Intelligent Data Analysis.* IDA-97, 1997, number 1280 in Lecture Notes in Computer Science, pp. 275–287, Springer.

Address for correspondence:
Erik Vullings
Delft Univ. of Technology, Dept. of Electrical Eng.
Mekelweg 4, P.O. Box 5031
2600 GA Delft, The Netherlands
E-mail: E.Vullings@et.tudelft.nl