Computer Methods and Programs in Biomedicine 68 (2002) 109–115

Computer Methods and Programs in Biomedicine

www.elsevier.com/locate/cmpb

Automatic pattern recognition in ECG time series

Karsten Sternickel

Institut für Strahlen-und Kernphysik, University of Bonn, Nussallee 14–16, D-53115 Bonn, Germany

Received 6 November 2000; received in revised form 9 March 2001; accepted 12 March 2001

Abstract

In this paper, a technique for the automatic detection of any recurrent pattern in ECG time series is introduced. The wavelet transform is used to obtain a multiresolution representation of some example patterns for signal structure extraction. Neural Networks are trained with the wavelet transformed templates providing an efficient detector even for temporally varying patterns within the complete time series. The method is also robust against offsets and stable for signal to noise ratios larger than one. Its reliability was tested on 60 Holter ECG recordings of patients at the Department of Cardiology (University of Bonn). Due to the convincing results and its fast implementation the method can easily be used in clinical medicine. In particular, it solves the problem of automatic P wave detection in Holter ECG recordings. © 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Pattern recognition; ECG; P wave; Wavelet transform; Neural networks PACS: 05.45.Tp; 42.30.Sy; 07.05.Mh

1. Introduction

One of the most important unsolved problems in ECG analysis is (despite intensive efforts) the automatic P wave detection within long-term Holter ECG recordings [1]. The P wave corresponds to the far field induced by a specific electrical phenomenon on the cardiac surface, namely atrial depolarization. Hence, knowledge about the morphology and occurrence of the P wave is essential for the investigation of the atrial electrophysiology [2–4]. The long-term observation, in particular, helps to characterize the underlying dynamics and is, therefore, necessary for the classification of atrial diseases. Numerous techniques ranging from digital filtering to linear and nonlinear methods [1] have been developed to recognize this specific waveform. Up to now, a reliable detection is still impossible, mainly because of the specific properties of the waveform: it has a small amplitude, it is generally deformed by offsets or contaminated with noise, its morphology has a temporal varying character, its distance to the well detectable QRS complex also varies in time, and it is represented by only a few datapoints (about 30 datapoints at a typical sampling rate of 200 Hz). Consequently, the P wave is the pattern most difficult to detect in ECG time series.

2. Method

Facing these problems, an approach for pattern recognition is suggested, basing on two parts: a feature extraction part and a classification part.

E-mail address: stern@iskp.uni-bonn.de (K. Sternickel).

PII: S0169-2607(01)00168-7

2.1. Feature extraction using the discrete wavelet transform (DWT)

The discrete wavelet transform [5,6] performs an adaptive time–frequency decomposition of a presented pattern. By the multiresolution representation it is possible to describe the signal structure by only a few coefficients in the wavelet domain Fig. 1.

Dilations and translations of a 'Mother function', or 'analyzing wavelet' define an orthogonal basis, the wavelet basis:

$$\Phi_{(s,l)}(x) = s^{-s/2}\,\Phi(2^{-s}x - l), \tag{1}$$

where s and l denote integers that scale and dilate the mother function $\Phi(x)$. The scale index s indicates the wavelet's width, and the location index l gives its position. Notice that the mother functions are rescaled, or 'dilated' by powers of two, and translated by integers. What makes wavelet bases especially interesting is the self-similarity caused by the scales and dilations. To span the data domain at different resolutions, the analyzing wavelet is used in a scaling equation:

$$W(x) = \sum_{k=-1}^{N-2} (-1)^k c_{k+1}\,\Phi(2x+k), \tag{2}$$

where $W(x)$ is the scaling function for the mother function, and c_k are the wavelet coefficients which must satisfy linear and quadratic constraints of the form

$$\sum_{k=0}^{N-1} c_k = 2, \qquad \sum_{k=0}^{N-1-2l} c_k c_{k+2l} = 2\delta_{l,0}, \tag{3}$$

where δ is the Kronecker delta. It is helpful to think of the coefficients $c_0, \ldots, c_n$ as a filter. The filter or coefficients are placed in a transformation matrix, which is applied to a raw data vector. The coefficients are ordered using two dominant patterns, one that works as a smoothing filter (like a moving average), and one pattern that works to bring out the data's 'detail' information. A more detailed description of the transformation matrix can be found in [7].

The wavelets used in this work are members of the Daubechies and Coiflet wavelet families, and the Haar wavelet [8]. The best performance was achieved by the Coiflet six wavelet because its structure adapts best to most P wave patterns.

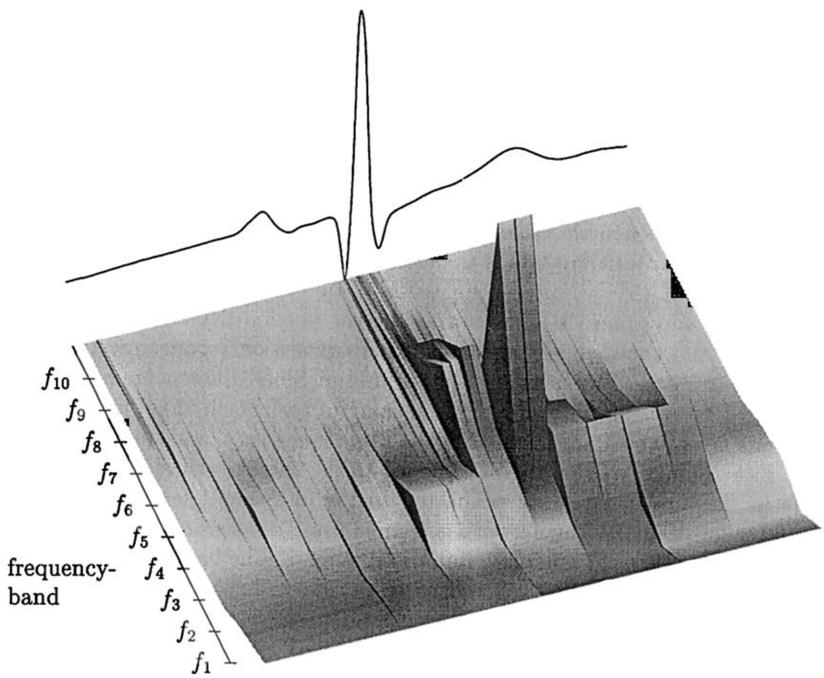

Fig. 1. Multiresolution representation of an ECG by wavelet coefficient. The values of the coefficients are translated into a three dimensional plot using a linear frequency scale.

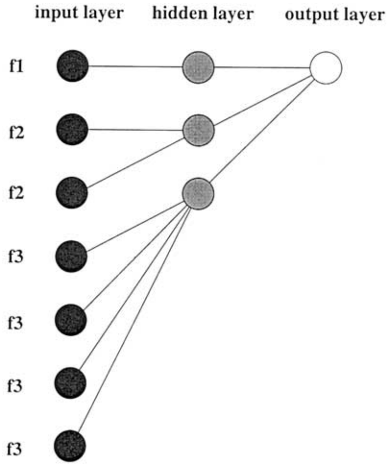

Fig. 2. Structure of a feed forward multi layer neural network using a frequency dependent connection. The neurons of the input layer are assigned to one single neuron in the hidden layer according to the frequency bands of the associated wavelet coefficients. The picture depicts seven input neurons out of three bands ($f1$–$f3$) connected to one of the three hidden neurons which are all linked to the single output neuron.

Table 1
Selected parameters for the example data set

Parameter	P wave	QRS complex
Number of selected patterns for the training set	10	10
Number of selected patterns for the validation set	20	20
Basis function	Coiflet 6	Haar
Block width for the DWT [datapoints]	32	16
Number of considered frequency bands	3	3
Detection probability to exceed for an event	0.8	0.6

2.2. Classification using Neural Networks

One of the major advantages of Neural Networks is their ability to generalize. This means a trained net could classify data from the same class as the learning data that it has never seen before. Therefore, the training patterns have to provide a sufficient characterization of the desired system behavior. In case of wavelet preprocessed P waves besides the network design two choices are important: the levels of course or fine detail representations and the amount of patterns used for the training. The wavelet coefficients of the lower frequency bands are decisive for the global shape of a P wave pattern. Those of higher frequency bands contain more information about details and are often proportionally stronger contaminated with noise which both would lead to a loss of generalization. Empirically, exclusion of the highest one or two frequency bands depending on the number of datapoints used for the wavelet transform (which is equivalent to the depth of frequency resolution) provided the best classification results. The amount of the training set is dependent on the variability of the shape of the P wave within the complete dataset and on the degree of contamination with noise. Generally, using commercial Holter systems and standard leads a range from 10 to 20 selected patterns out of a exemplary 10 min sequence is necessary for a sufficient generalization. In case of strong noise it is advantageous to select more patterns from different sequences. In order to create an additional validation set, some more P wave patterns (randomly distributed over the time series) have to be selected.

The chosen set of coefficients is assigned to a three layer feed forward Neural Network [9–11] whereby each coefficient represents a neuron in the input layer. The choice of the dimension of the hidden layer is the most crucial point during the design of the network topology. It has to be large enough to solve the learning problem and small enough to prevent overadaption. A suitable solution is to correlate the dimension with the number of considered frequency bands according to the coefficients in the input layer. That means, if n bands are considered the dimension should be n also so that each neuron in the hidden layer is assigned to exactly one band. The output layer has only one neuron since the learning task is to decide between an event and no event. Then, the value of the output neuron tends to unity in case of a detected P wave and to zero otherwise. There are different possibilities to link a network. If no initial knowledge is available a full connection (connection from each layer to its succeeding

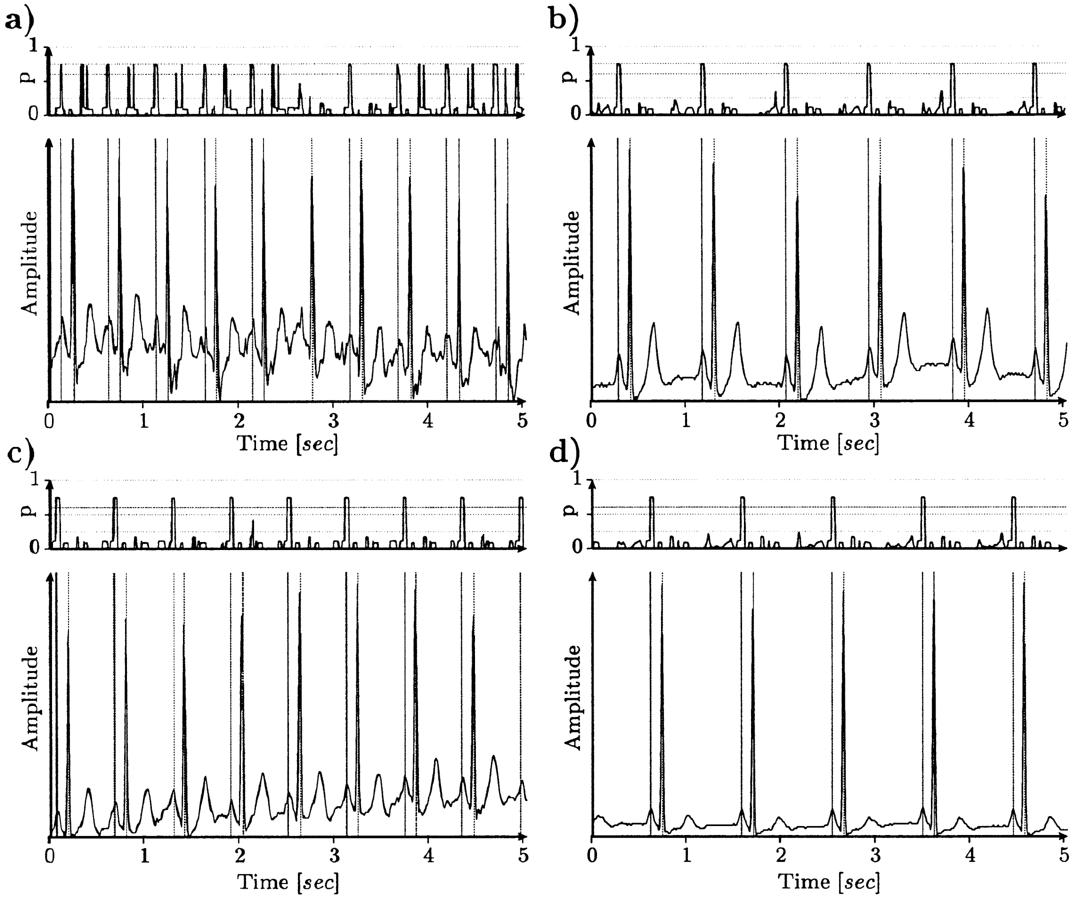

Fig. 3. P wave detection in an ECG time series. The lower pictures show segments of an ECG of a healthy subject taken in the temporal distance of 4 h. All QRS complexes are marked before the actual P wave recognition. P waves are marked only if the probability P (upper pictures) exceeds the cutoff level (horizontal line). The detection is even reliable for a varying morphology of the P wave although the amount of FP increases in case of contamination with noise (a).

layer) is a usual linkage. Since different frequency bands should be weighted equally it is meaningful to link all these neurons of the input layer, which are member of the same frequency band, to the one neuron in the hidden layer, which is assigned to the corresponding band. All neurons in the hidden layer are connected to the one single neuron in the output layer (see Fig. 2). The network is trained using a standard backpropagation algorithm [12] with randomly initialized weights. The backpropagation weight update rule, also called *generalized delta rule* reads as follows:

$$\Delta w_{ij} = \eta \delta_j \, o_i. \tag{4}$$

δ_j

Table 2
Classification result of the example dataset. Additionally, the results for two other wavelets are indicated. In this case the Coiflet 6 is most suitable for the P wave detection. Using the other wavelets much more FP are needed to reach a sufficient amount of TP

	Coiflet 6	Daubechies 12	Haar
$N_{\mathrm{P}}^{\max}$	112 897	112 897	112 897
N_{event}	138 699	253 201	280 498
TP	112 808 (99.9%)	103 874 (92.0%)	112 478 (99.6%)
FP	28 942	149 327	168 020

$$\delta_j = \begin{cases} f'_j(\text{net}_j)(\text{t}_j - \text{o}_j) & \text{if neuron } j \text{ is an output neuron} \\ f'_j(\text{net}_j) \sum_k \delta_k w_{jk} & \text{if neuron } j \text{ is a hidden neuron} \end{cases} \tag{5}$$

where η is the learning factor (a constant), δ_j the error of neuron j (the difference between the real output and the teaching input), net_j the net input in neuron j, t_j the teaching input of neuron j, o_i the output of the preceding neuron j, i the index of a predecessor of the current neuron j with link w_{ij} from i to j, j the index of the current neuron, and k the index of a successor to the current neuron j with link w_{jk} from j to k. A new activation $a_j(t)$ of neuron j in step t is computed using the sigmoidal function as transfer function f:

$$a_j(t+1) = f_{\text{act}}(\text{net}_j(t), a_j(t), \theta_j), \tag{6}$$

with

$$f_{\text{act}}(x) = \frac{1}{1+\text{e}^{-x}}, \tag{7}$$

where θ_j is the threshold (bias) of neuron j. The new activation at step t lies in the range 0–1. The bias determines where the activation function has its steepest ascent and is treated like a weight during training. To compute new activation values for the neurons all of them have to be visited in a special order. Here, a topological order is chosen (the neurons are sorted by their topology). This order corresponds to the natural propagation of activity from input to output.

Backpropagation is a supervised technique of training, which means that training coincides in minimizing an error function. Generally, this procedure is critical because a local minimum can be found whereas an absolute minimum is required for the learning function. However, to reach best generalization the learning should be stopped in the minimum of the validation set error. When learning is not stopped, overtraining occurs and the performance of the net on the whole data set decreases, despite the fact that the error on the training data still gets smaller. Additionally, for a better discrimination it is useful to arbitrarily add some (transformed) templates out of the time series containing anything but P waves to the training set and to shuffle all patterns between the single learning steps. The training procedure is repeated several times until the net reaches a sufficient performance. The trained and tested net is finally linked to an executable program. After this, the complete time series is analyzed using the moving window technique. A window of the chosen wavelet width is moved over the time series in steps of one data point. The datapoints of the actual window are wavelet transformed and valued by the trained network. This generates a new time series consisting of probabilities for a P wave event. Strong variations of the P wave may lead to a decreased detection probability whereas the separation from other events is still maintained. Hence, a patient dependent cutoff value provides the best P wave extraction. Nevertheless, the knowledge about the principle appearance within the time series helps to avoid contingent detection errors. Due to the fact, that P waves always appear within a certain range before a QRS complex an additional selective criterion is available. Especially, in case the time series is strongly contaminated by artifacts it is favorably to maximize the detecting sensitivity at the expense of specificity, i.e. by decreasing the number of considered frequency bands.

2.3. Example

The long-term ECG of a healthy subject is analyzed using raw ECG signals recorded by a commercial Holter system (Elatec, ELA Medical) and digitized with 200 Hz. The dataset has an average recording quality without larger series of artifacts. Generally, due to its amplitude and its characteristic shape the QRS complex is a well detectable signal. The presented method can be applied easily to this signal form using suitable parameters. Table 1 summarizes the selected parameters for the wavelet transform and the net learning process for P waves and QRS complexes of the example dataset.

Because a long-term ECG contains on average more than 100 000 events, it is impossible to control the success of the method manually. To define a maximum number of possible P waves (N_P^{max}) within a complete time series N_P^{max} is set the amount of all recognized QRS complexes of normal heart beats (N_{QRS}). Now, a correct classification (CC) is determined by a detected event within a certain range before an according QRS complex. The number of true positives (TP) is the total amount of all CC and the number of false positives (FP) is set to be $FP = N_{event} - TP$ where N_{event} is the total amount of discovered events. Fig. 3 contains four sequences out of the complete time series: (a) biking; (b) recreation phase; (c) car driving; (d) sleep demonstrating the strong variability of the P wave morphology during a long-term observation. Only in case of noise contamination, e.g. by movement artifacts as visible in Fig. 3a the amount of FP increases, whereas the true positions are still discovered. The results for this dataset are listed in Table 2. Comparatively, the results for two other wavelets are added. The total quantity of normal heartbeats (112 897) detected by the described procedure lies close to the value reported by the commercial software (ELATEC V3.03 B), which counted 112 753 beats.

3. Results

An executable program was developed for the method outlined above and installed at the Department of Cardiology (University of Bonn). Long-term ECGs of 60 patients with intermittent atrial fibrillation and of 20 healthy subjects all recorded with Elatec's SYNESIS recorder and digitized with 200 Hz were analyzed. Each dataset was processed using the same parameters as listed in Table 1 for the QRS and P wave detection. In all cases with sinusrhythm the total amount of detected P waves was greater than 92% in relation to the located QRS complexes (range: 92.68–99.99%, mean: 96.19%). The required time for the preprocessing of QRS complexes and P waves (marking, wavelet transform, neural network training) for one 24 h data set takes about 10 min. A Pentium III Computer with 500 MHz needs about 5 min computing time for the moving window procedure. Of course, a complete automation without any manual preprocession can easily be performed.

4. Conclusion

Due to its reliable detecting quality and the possibility of fast and easy implementation the proposed method is a suitable tool for automatic pattern recognition in ECG time series, particularly in clinical routine. Although a similar method [13] has been developed for the automated detection of late potentials it has not achieved convincing results because only one wavelet has been used and the frequency bandwidth has not been optimally adjusted. That shows the importance of a suitable choice of the wavelet basis function and an optimal design of the network.

Acknowledgements

The author thanks Professor Dr P. David for his permanent support and encouragement. Also, the author very much appreciates the support by Professor Dr B. Lüderitz (Department of Cardiology, University of Bonn) beginning with the ECG measurements of patients and continuing during the fruitful collaboration with his staff Professor Dr W. Jung, Dr L. Lickfett, Dr C. Wolpert, and Dr A. Bitzen.

References

[1] N. Maglaveras, T. Stamkopoulos, K. Diamantaras, C. Pappas, M. Strintzis, ECG pattern recognition and classification using non-linear transforms and neural networks: a review, Int. J. Med. Inform. 52 (1998) 191–208.

[2] P.E. Dilaveris, E.J. Gialafos, G.K. Andrikopoulos, D.J. Richter, V. Papanikolaou, K. Poralis, Clinical and electrocardiographic predictors of recurrent atrial fibrillation, Pace-Pacing Clin. Electrophysiol. 23 (3) (2000) 352–358.

[3] P. Stafford, P. Denbigh, R. Vincent, Frequency analysis of the P wave: comparitive techniques, Pacing Clin. Electrophysiol. 18 (2) (1995) 261–270.

[4] S. Ciaroni, L. Cuenoud, A. Bloch, Clinical study to investigate the predictive parameters for the onset of atrial fibrillation in patients with essential hypertension, Am. Heart J. 139 (5) (2000) 814–819.
[5] S.G. Mallat, Multiresolution approximation and wavelet orthonormal bases of L2, IEEE Trans. Am. Math. Soc. 315 (1989) 69–87.
[6] A. Graps, An introduction to wavelets, IEEE Comput. Sci. Eng. 2 (2) (1995) 50–61.
[7] W. Press, et al., Numerical Recipes in C, Cambridge University Press, New York, 1992, pp. 591–606.
[8] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pat . Rec. Mach. Intel. 11 (7) (1989) 674–693 Table 1.
[9] S.S. Cross, R.F. Harrison, R.L. Kennedy, Introduction to neural networks, Lancet 346 (1995) 1075–1079.
[10] M. Peltorante, G. Pfurtscheller, Neuronal Network Based Classification of Single-trial Data, Artificial Intelligence in Medicine, vol. 5, Elsevier Science Publisher, 1993, pp. 503–513.
[11] A. Zell, G. Mamier, M. Vogt, et al, SNNS (Stuttgart Neuronal Network Simulator), Institute for Parallel and Distributed High Performance Systems (IPVR), University of Stuttgart, Germany, 1994.
[12] D. Hush, B. Horne, Progress in supervised Neural Networks: what's new since Lippman?, IEEE SP Mag. 1 (1993) 8–39.
[13] A. Rakotomamonjy, B. Migeon, P. Marche, Automated neural network detection of wavelet preprocessed electrocardiogram late potentials, Med. Biol. Eng. Comput. 36 (3) (1998) 346–350.