

Clustering ECG Complexes Using Hermite Functions and Self-Organizing Maps

Martin Lagerholm, Carsten Peterson*, Guido Braccini, Lars Edenbrandt, and Leif Sörnmo, *Member, IEEE*

Abstract—An integrated method for clustering of QRS complexes is presented which includes basis function representation and self-organizing neural networks (NN's). Each QRS complex is decomposed into Hermite basis functions and the resulting coefficients and width parameter are used to represent the complex. By means of this representation, unsupervised self-organizing NN's are employed to cluster the data into 25 groups. Using the MIT-BIH arrhythmia database, the resulting clusters are found to exhibit a very low degree of misclassification (1.5%). The integrated method outperforms, on the MIT-BIH database, both a published supervised learning method as well as a conventional template cross-correlation clustering method.

Index Terms—Clustering, Hermite functions, QRS complex, self-organizing networks.

I. INTRODUCTION

COMPUTER-BASED methods for analysis and interpretation of electrocardiograms (ECG's) have been subject to intense research for nearly four decades. Electrocardiographs that perform sophisticated signal processing and diagnostic interpretation of signals are today wide-spread [1]. Artificial neural networks (ANN's) are one of the most recent techniques in this field and have primarily been considered for classification of ECG's into different diagnostic groups [1], [2]. It has been shown that ANN's for specific issues can perform better than both experienced cardiologists and ruled-based criteria, e.g., in detecting acute myocardial infarction from the ECG [3]. A first generation of ANN's have also been implemented in commercial electrocardiographs [4].

Visual analysis of long-term (24 h) ECG is tedious and operator dependent. Computer techniques have been developed in order to facilitate visual analysis, e.g., by condensed print-outs of various signals and trends. With this type of presentation the operator usually can analyze a 24-h recording in 20–40 min provided that no complex arrhythmias exist. It is obvious that automated systems for detection of arrhythmias considerably re-

duces the amount of time the operator needs to spend. Several commercial systems are available for long-term ECG analysis. However, their performance deteriorate markedly when noise and artifacts are present and, as a consequence, an excessive number of beat classes is created which require considerable manual editing.

Beat clustering based on ANN requires that a set of descriptive measurements first is extracted from the ECG. The most commonly used approaches taken to this problem is to use a set of heuristic features e.g., measurements of QRS wave amplitudes and durations [5], or time samples selected from the QRS interval [6]–[8]. While these approaches are likely to perform well in recordings acquired during favorable conditions, it is well-known that classification based on heuristic features or the unprocessed signal itself are more vulnerable to noise than are basis function representations [9]–[11]. Electrocardiographic feature extraction by basis function representation was suggested already in the 1960's when Laguerre orthogonal functions were proposed [12]. Later, the Karhunen–Loeve (KL) expansion, which provides an optimal signal representation in the mean square error sense, was found to be suitable for this purpose, e.g., [13] and [14]. It is well-known that the KL basis functions constitute an orthonormal set and, therefore, each coefficient in the expansion represents independent information. Inspired by the shape of the most significant KL basis functions, the orthogonal Hermite functions were later proposed as a useful parametric model of the QRS complex [15]. The main advantage with the Hermite model is that it includes a width parameter which provides an efficient representation of beats with large differences in QRS duration, e.g., normal beats and PVC's; the KL approach cannot as easily handle such differences. The Hermite model was later studied within the context of real-time monitoring [16] with application to e.g., data compression [17]. In that study, a multiple-input adaptive linear combiner was developed for the purpose of jointly estimating the Hermite model parameters, including the width.

In this paper, we develop a method for unsupervised characterization of ECG signals. The aim is to partition the beats into clusters that represent central features of the data such that similarity structures between clusters are preserved. Our approach involves Hermite function representation of the QRS complexes in combination with self-organized clustering. In contrast to feedforward supervised ANN learning algorithms, which are trained to identify predefined features in the data, self-organizing networks (SOM) organize themselves according to the natural structure of the data—no external teacher is called for [18]. Here, we use a variant of an SOM where neighbors are updated according to a soft-max (Potts) prescription. The input

Manuscript received February 17, 1999; revised March 3, 2000. This work was supported by the Swedish National Board for Technical Development and by the Swedish Medical Research Council. Asterisk indicates corresponding author.

M. Lagerholm is with the Complex Systems Group, Department of Theoretical Physics, Lund University, Sölvegatan 14A, S-223 62 Lund, Sweden.

*C. Peterson is with the Complex Systems Group, Department of Theoretical Physics, Lund University, Sölvegatan 14A, S-223 62 Lund, Sweden (e-mail: carsten.peterson@thep.lu.se).

G. Braccini and L. Edenbrandt are with the Department of Clinical Physiology, Lund University, S-221 85 Lund, Sweden.

L. Sörnmo is with the Signal Processing Group, Department of Applied Electronics, Lund University, Box 118, S-221 00 Lund, Sweden.

Publisher Item Identifier S 0018-9294(00)05122-3.

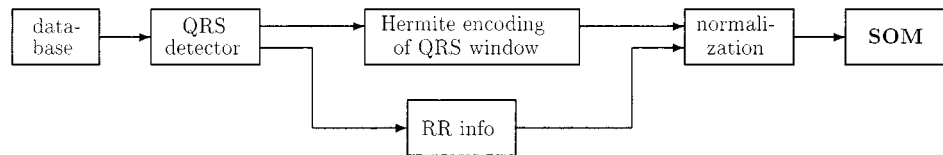


Fig. 1. A flow-chart of our method.

data vector consists of the Hermite coefficients, the width parameter and two measures related to the length of the RR interval. It should be stressed that the primary objective is to perform clustering; the final beat classification is performed by an expert.

Both our key algorithmic ingredients, Hermite function representation of the beats and self-organizing networks for the clustering, represent powerful generalizations of more conventional approaches, KL basis representations and standard K-means clustering, respectively. These methods, which each can be viewed as special cases from our approaches, in general have inferior performance. Hence, we did not explore these less powerful methods on factorized parts of the problem. In Fig. 1 a flow-chart of our integrated method with its different components is shown.

Recently, an ANN-based method for beat classification was presented, which uses the “mixture-of-expert” (MOE) principle [19]. The present work differs distinctly in several respects to that work, e.g., whereas our approach handles all steps from QRS detection to beat clustering, the work in [19] is focused on patient-adaptable techniques that require initial labeling of beats. To the extent it is feasible, our results are compared to those in [19] using the MIT-BIH Arrhythmia database [20].

This paper is organized as follows. In Section II, we briefly describe the MIT-BIH arrhythmia database. Our preprocessing technique based on the Hermite function decomposition is developed in Section III. Section IV describes the essentials of the self-organizing map and Section V presents the results and comparisons with a mixture-of-expert model and a correlation-based method. Finally, a discussion and conclusions are found in Section VI.

II. ECG DATABASE

A total of 48 ECG recordings were studied from the MIT-BIH arrhythmia database [20]. Each recording has a duration of 30 min and includes two leads, denoted A and B, respectively, which are chosen among the modified limb lead II and the modified leads V1, V2, V4, or V5. All beats in the database were annotated by two cardiologists: approximately 70% of the beats were classified as normals while the remaining types of beats were divided into 16 different kinds of abnormal beats. Since we focus on the QRS part of the signal, the “p” beats only containing the P wave are omitted. Hence, we only use 15 abnormal beat types (see Table I). When comparing our procedure to MOE [19], we utilize all the 15 different categories rather than lumping them together as normal/abnormal beats.

The ECG signals were resampled from the original 360-Hz sampling rate to 1 kHz in order to be compatible with our existing software for high-resolution ECG analysis, including an

TABLE I
THE 16 DIFFERENT TYPES OF BEATS FROM THE MIT-BIH ARRHYTHMIA DATABASE WITH ANNOTATIONS. ALSO SHOWN IS HOW THESE ARE IDENTIFIED WITH “OUR METHOD” (THE ECG SIGNAL-DETECTION METHOD DESCRIBED IN THE TEXT) AND THE DIFFERENCES (Δ) WITH THE DATABASE ANNOTATION. THE POOR DETECTION RATE OF THE a-BEATS WAS DUE TO THE FACT THAT SEVERAL VERY LOW AMPLITUDE PVC’S IN RECORD 201 WERE MISSED BY THE QRS DETECTOR

code	beat type	Annotation	Our Method	Δ (%)
N	Normal beat	75053	74854	0.3
L	Left bundle branch block	8074	8060	0.2
R	Right bundle branch block	7259	7239	0.3
A	Atrial premature beat	2544	2538	0.2
a	Abberated atrial premature beat	150	109	27.0
J	Nodal (junctional) premature beat	83	83	0.0
S	Supraventricular premature beat	2	2	0.0
V	Ventricular premature beat	7129	7110	0.3
F	Fusion of ventricular and normal beat	803	798	0.6
b	Ventricular flutter wave	472	458	0.3
e	Atrial escape beat	16	16	0.0
j	Nodal (junctional) escape beat	229	228	0.4
E	Ventricular escape beat	106	106	0.0
P	Paced beat	7028	7014	0.2
f	Fusion of paced and normal beat	982	979	0.3
Q	Unclassifiable beat	33	33	0.0

algorithm for QRS detection [21]. The resampling was done by functions available in the Matlab software (Mathworks Inc., Nattick, MA) for which suitable interpolation and decimation factors were selected. Obviously, this increase in sampling rate does not affect the spectral content of the signal but implies that a fiducial point with better time resolution can be determined for the QRS morphology.

III. PREPROCESSING

The coefficients of the Hermite basis function representation are determined from a high-pass filtered ECG signal in order to avoid poor fits due to the presence of very low-frequency components. The baseline filter was implemented as a linear-phase, finite impulse response low-pass filter which estimates the baseline wander followed by subtraction of this estimate from the original ECG. The filter complies with the American Heart Association (AHA) recommendations in terms of cutoff frequency for baseline filtering [22]. The fiducial point of each detected

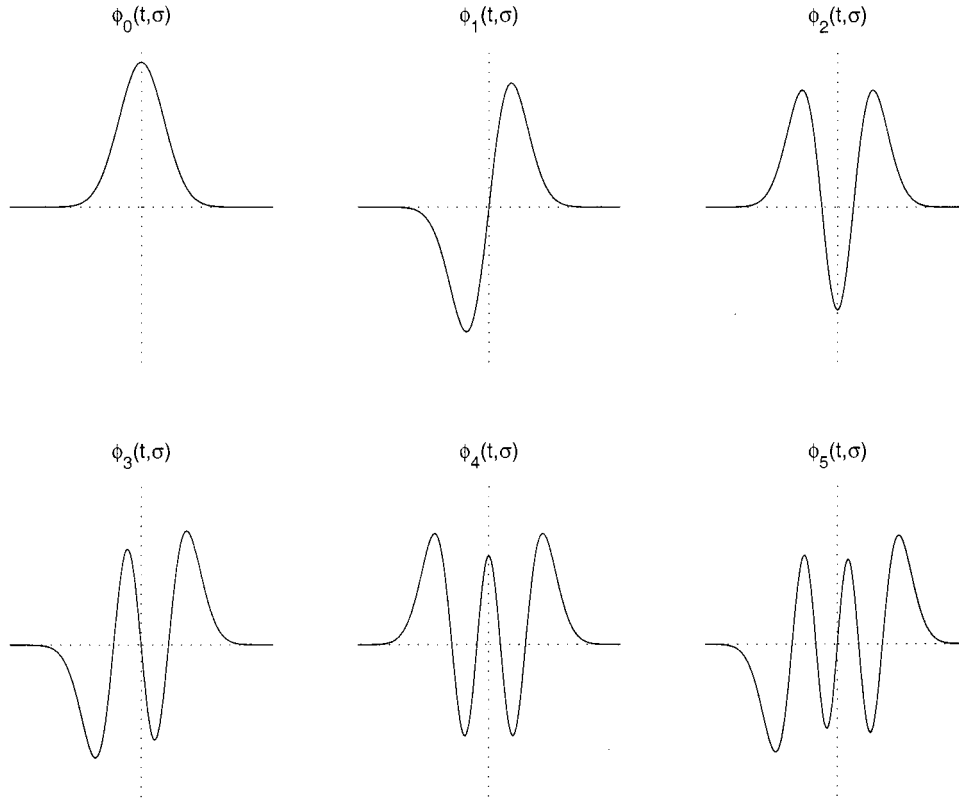


Fig. 2. The first six Hermite functions (2, 3) plotted as a functions of t (same scale in all figures) and for the same σ .

QRS complex was defined by the peak location in an envelope signal obtained by summing the envelopes of each individual lead; further details on the envelope computation can be found in [21]. Next, the ECG signal is extracted from a $t_0 = 200$ -ms window centered around the fiducial point. The window length is selected to ensure that also very wide duration PVC's were included in the analysis window. A window of size 150 ms would cover the QRS complex of normal beats but are not long enough for all the types of artifacts, 250 ms seems to be unnecessarily long. The overall accuracy of this detect-extract QRS-detection procedure when comparing with the database annotation file is high -99.7% . The 0.3% discrepancy is broken down into beat types in Table I.

Each QRS complex is represented by a $2t_0$ -dimensional vector $\vec{x} = \{x_t\}$. The extra t_0 ms is used to enforce that the encoded beats are close to zero outside the QRS complex. This is done by adding a 100-ms zero signal on each side of the original beat. The QRS complexes \vec{x} are expanded onto a linear combination of Hermite basis

$$x_t = \sum_{n=0}^{N-1} c_n(\sigma) \phi_n(t, \sigma) + e(t, \sigma) \quad (1)$$

where

$$\phi_n(t, \sigma) = \frac{1}{\sqrt{\sigma 2^n n! \sqrt{\pi}}} e^{-t^2/2\sigma^2} H_n(t/\sigma) \quad (2)$$

and $H_n(t/\sigma)$ are the Hermite polynomials. The width σ approximates the half-power duration. With $H_0(x) = 1$ and $H_1(x) = 2x$, the Hermite polynomials are recursively given by

$$H_n(x) = 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x). \quad (3)$$

Hence, each complex is represented by $N + 1$ parameters; σ and $c_n(\sigma)$ ($n = 0, 1, \dots, N-1$). The first six Hermite functions are shown in Fig. 2. The Hermite functions form an orthonormal basis

$$\sum_{t=-\infty}^{\infty} \phi_n(t, \sigma) \phi_m(t, \sigma) = \delta_{mn} \quad (4)$$

for any fixed value of σ , which allows for a cost-effective calculation of $c_n(\sigma)$ in x_t in (1).

For a finite window size, $\phi_n(\sigma)$ does not obey (4). However, if $\phi_n(\sigma)$ is close to zero outside the window it is still a good approximation. In particular, this problem occurs for large n and σ , (see Fig. 2); for a given window size, N imposes an upper limit σ_0 on σ . On the other hand, for a good representation of x_t , σ_0 should not be too small. Fortunately, fairly small values of N , i.e., 3 or 4, allow for good beat representation. A window size of $2t_0 = 400$ ms, including the 100-ms zero padding on each side of the QRS complex, is in this respect a suitable choice. Defining $\phi_n(\sigma)$ as being close enough to zero outside the window if

$$|\phi_n(-t_0, \sigma)| = |\phi_n(t_0, \sigma)| < \frac{1}{10} \max_{t \in [-t_0, t_0]} |\phi_n(t, \sigma)| \quad (5)$$

and

$$|\phi_n(t, \sigma)| \leq |\phi_n(t_0, \sigma)| \quad \text{for all } |t| > t_0 \quad (6)$$

one obtains for $N = 3, 4, 5$, and 6 the maximum widths $\sigma_0 = 62, 55, 51$, and 47 ms, respectively. These numbers are sufficiently large for good representation of the beats as is evident from Table II.

TABLE II

THE RELATIVE ERROR ϵ AND THE AVERAGE WIDTH OF THE HERMITE FUNCTION, $\langle\sigma\rangle$ AVERAGED FOR ALL BEATS AND OVER BOTH LEADS. RESULTS SHOWN ARE FOR $N = 3, 4, 5$ AND 6 HERMITE FUNCTIONS, RESPECTIVELY. ALSO SHOWN ARE THE MAXIMAL ALLOWED WIDTHS σ_0 ACCORDING TO (5) AND (6)

Hermite functions (N)	3	4	5	6
ϵ (%)	9.7	6.8	5.5	4.5
$\langle\sigma\rangle$ (ms)	19.1	19.0	18.7	18.4
σ_0 (ms)	62	55	51	47

For a given σ , the coefficients $c_n(\sigma)$ are determined by minimizing the summed square error

$$\sum_t |e(t, \sigma)|^2 = \sum_t \left| x_t - \sum_n c_n(\sigma) \phi_n(t, \sigma) \right|^2. \quad (7)$$

The minimum of (7), for a fix σ , is conveniently evaluated by means of the orthogonality property (4), given by

$$c_n(\sigma) = \vec{x} \cdot \vec{\phi}_n(\sigma). \quad (8)$$

The optimal combination of σ and the coefficients is determined by stepwise increments of σ up to σ_0 and then recomputation of (8) rather than fitting all parameters simultaneously in order to have a fast decomposition. The use of (8) significantly reduces the computational demand when decomposing the signals in Hermite functions. This represents an asset in real-time clustering.

As a relative error measure we use

$$\epsilon = \frac{\vec{e}(\sigma)^2}{\vec{x}^2}. \quad (9)$$

In general, beats are well represented even with a small number of Hermite functions. This is confirmed both by visual inspection and by the fact that ϵ is low (see Table II). The P and f beat types, which are pacemaker related, and Q (unclassifiable) are less suitable for representing with a few Hermite functions. However, the variance of ϵ between different types of beats is clearly acceptable when five Hermite functions are used (see Table III). Fig. 3 exemplifies the representation of a normal beat with 3, 4, 5, and 6 Hermite functions.

Apart from the above parameters related to the Hermite function representation, local rhythm information will also be used as input to the SOM. The following two features are considered:

$$\begin{aligned} R^- &\equiv R_i^- = R_i^* - R_{i-1}^* \\ R^+ &\equiv \Psi [(R_{i+1}^- - R_i^-) - (R_i^- - R_{i-1}^-)] \end{aligned} \quad (10)$$

where R_i^* is the occurrence time for the i ; the beat, $\Psi(x) = x$ if $x > 0$, otherwise $\Psi(x) = 0$. In this way R^+ does not depend on the heart rate, because it is significantly different from zero only if the i :th beat is premature. This is an important feature for atrial beats.

TABLE III

THE RELATIVE ERROR ϵ INTRODUCED WHEN DECOMPOSING THE BEATS INTO FIVE HERMITE BASIS FUNCTIONS. THE NUMBERS ARE AVERAGED OVER BOTH LEADS (A AND B) AND ALL BEATS IN THE DATABASE

	N	L	R	A	a	J	S	V	F	b	e	j	E	P	f	Q
ϵ (%)	5	2	8	9	7	5	4	5	3	6	3	4	4	14	11	19

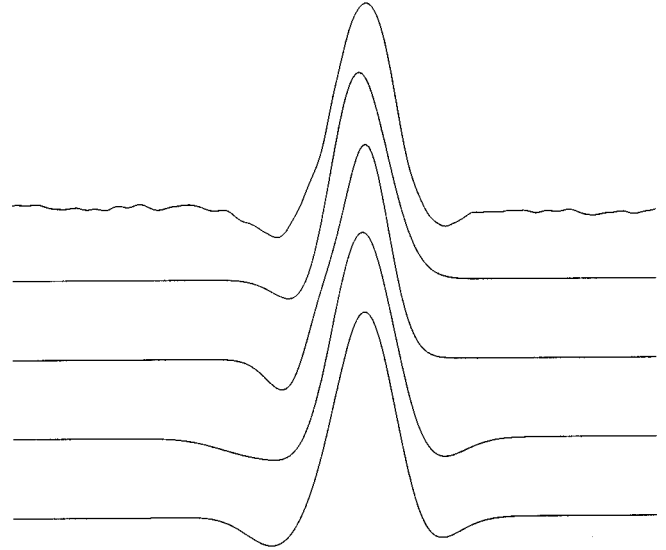


Fig. 3. An example of Hermite function representation. The upper curve is the original beat (lead A in beat #1 of record 100). The subsequent four curves represent decomposition using 3, 4, 5, and 6 Hermite functions, respectively.

IV. SELF-ORGANIZING MAP

A. Self-Organizing Networks

Whereas feedforward supervised ANN learning algorithms are aimed at identifying predefined features in the data, SOM's operate in a different way. The desired feature values are not known beforehand and the network must organize itself according to the natural structure of the data. Hence, this approach falls within the family of clustering methods.

1) *Topological Maps*: The idea is to map a set of K -dimensional input vectors $\vec{\xi}$ onto a discrete two-dimensional (2-D) space with M positions (see Fig. 4). Each position in the output space is represented by an output node, h_j ($j = 1, \dots, M$). For each h_j , a position in the input space is associated by $\vec{\omega}_j$. The distance between h_j and input vector $\vec{\xi}$ is defined as

$$d_j = \sum_{k=1}^K (\omega_{jk} - \xi_k)^2 \quad (11)$$

where the index labeling the different input vectors is suppressed. In cases, where each input vector is normalized (each input, ξ_k , over all the beats) to zero mean and one, and unit variance, this simple measure is sufficient. For each input vector $\vec{\xi}$, a "winner" h_{i^*} is chosen as the one closest to $\vec{\xi}$, i.e.,

$$d_{i^*} = \min_j d_j. \quad (12)$$

Input vectors having h_{i^*} as a winner are said to "belong" to cluster (output node) i^* . In what follows, a cluster will also be

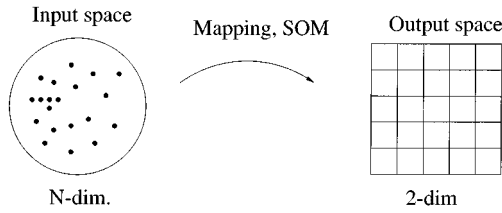


Fig. 4. The SOM mapping from a high-dimensional space to a discrete 2-D space.

referred to as a *box*. The positions of h_j in the input space \vec{w}_j could be viewed as a weight in the mapping process. This is the commonly used terminology in the ANN literature. In the simplest form of self-organization, “winner-takes-all,” the weights ω_{mk} are updated such that h_m will become more sensitive to the presented pattern, similar to standard clustering algorithms. This is accomplished by

$$\omega_{mk} \rightarrow \omega_{mk} + \Delta\omega_{mk}$$

$$\Delta\omega_{mk} = \begin{cases} \eta(\xi_k - \omega_{mk}), & \text{if } m = i^* \\ 0, & \text{if } m \neq i^* \end{cases} \quad (13)$$

where η is the step size toward the input vector, (parameter changing) which is changed during training.

It is often desirable to have a topological order among the output units h_j , such that neighboring units in the output space will respond to adjacent points in the input space. A commonly used technique to achieve such a *topological map* is by updating units close to the winner node in the same way as the winner node [18] but with a step that in addition to η also depends upon the distance to the winner, r (in the output space). In this way neighboring units will end up with similar weight vectors. We do this using *Potts* neurons, where the winner “shares” its step with its neighbors. The Potts neuron, \vec{v}_{i^*j} , encodes the relative strengths according to

$$v_{i^*m} = \frac{e^{-r_{i^*m}^2/T}}{\sum_k e^{-r_{i^*k}^2/T}} \quad (14)$$

The distance r_{ij} (in the output space) between the output units defines the topology of the network. The artificial “temperature” or width T is decreased (annealed) as learning proceeds. The update rule (13) is then modified to

$$\Delta\omega_{mk} = \eta(\xi_k - \omega_{mk})v_{i^*m} \quad \text{for all } m. \quad (15)$$

Note that

$$\sum_j v_{ij} = 1 \quad (16)$$

expressing the “share” philosophy. For $T \rightarrow 0$ (15) reduces to the “winner-take-all” update rule (13). Suitable parameter choices for our procedure are found below.

B. The QRS Map

1) *Input Structure*: We map our preprocessed QRS data, denoted by ξ_k , onto 5×5 maps using (11) and (12). The ξ_k -representation is shown in Table IV where c_i and σ refer to the Hermite decomposition for each lead. Since the components of the input vector (c_n, σ, R^- and R^+) have different origins one does

TABLE IV
DEFINITION OF THE SOM INPUT COMPONENTS ($\xi_1 \dots \xi_{14}$) WHEN USING FIVE HERMITE FUNCTIONS: c_k AND σ DENOTE COEFFICIENTS AND WIDTHS IN THE HERMITE FUNCTION DECOMPOSITION AND R^- AND R^+ ARE DEFINED IN (10)

	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6	ξ_7	ξ_8	ξ_9	ξ_{10}	ξ_{11}	ξ_{12}	ξ_{13}	ξ_{14}
lead A	c_1	c_2	c_3	c_4	c_5	σ								
lead B							c_1	c_2	c_3	c_4	c_5	σ		
lead A/B													R^-	R^+

not *a priori* know their relative importance for distinguishing between various types of beats. Hence, we transform the components by subtracting the average and scaling to a common unit variance.

We have a mapping from a 12-dimensional input space to a 2-D (5×5) output space. This is in contrast to many applications using SOM’s, where the key goal is to reduce the dimensionality of the problem. In our case, that reduction has been achieved with the Hermite function decomposition. One central reason here for using SOM’s is to benefit from its topological structure when interpreting the data.

2) *Training*: The training, which is divided into two phases, requires setting a few parameters. However, these turn out to be rather insensitive, i.e., no fine tuning is needed. In phase I, the output vectors evolve with a strong neighborhood dependence, whereas phase II proceeds with a weaker dependence. The latter allows for outliers to be identified, which are important for diagnosis. Below we describe in more detail the parameters and procedures we use.

a) *Initializing the weights*: The input data typically occupies a very small subspace of the total input space. Ideally, one wants the output nodes to focus on this subspace from the start. Since the input vector components have been normalized to zero mean, this is accomplished initializing ω_{jk} using a uniform distribution on the interval $[-\varepsilon, \varepsilon]$, where ε is chosen small as compared to the standard deviation ($\varepsilon = 0.01$).

b) *Local neighborhood*: In order to produce a good topological map, one should start with a wide neighborhood (T large), compared to the size of the network. This will force all units to become more or less similar. This will force all the cluster centers to the same region of the input space. The neighborhood is then decreased during training, allowing the units to focus on local regions of the input space. For a too small T , the coupling between the units is broken and no ordering will emerge. On the other hand, if T is large all the time, all units in the network will end up pointing in the same direction. We decrease the neighborhood according to

$$T_{t+1} = cT_t \quad (17)$$

where t is the iteration number and $c = 0.95$. The initial T is estimated from the size of the output map as

$$T_0 = -\frac{d^2}{\ln(2)} \quad (18)$$

where d denotes the diagonal distance of the map (with square boxes with unit length and a 5×5 map, d would be $5\sqrt{2}$). This implies that a output vector at distance d from the winner will be updated with half the strength of the winner.

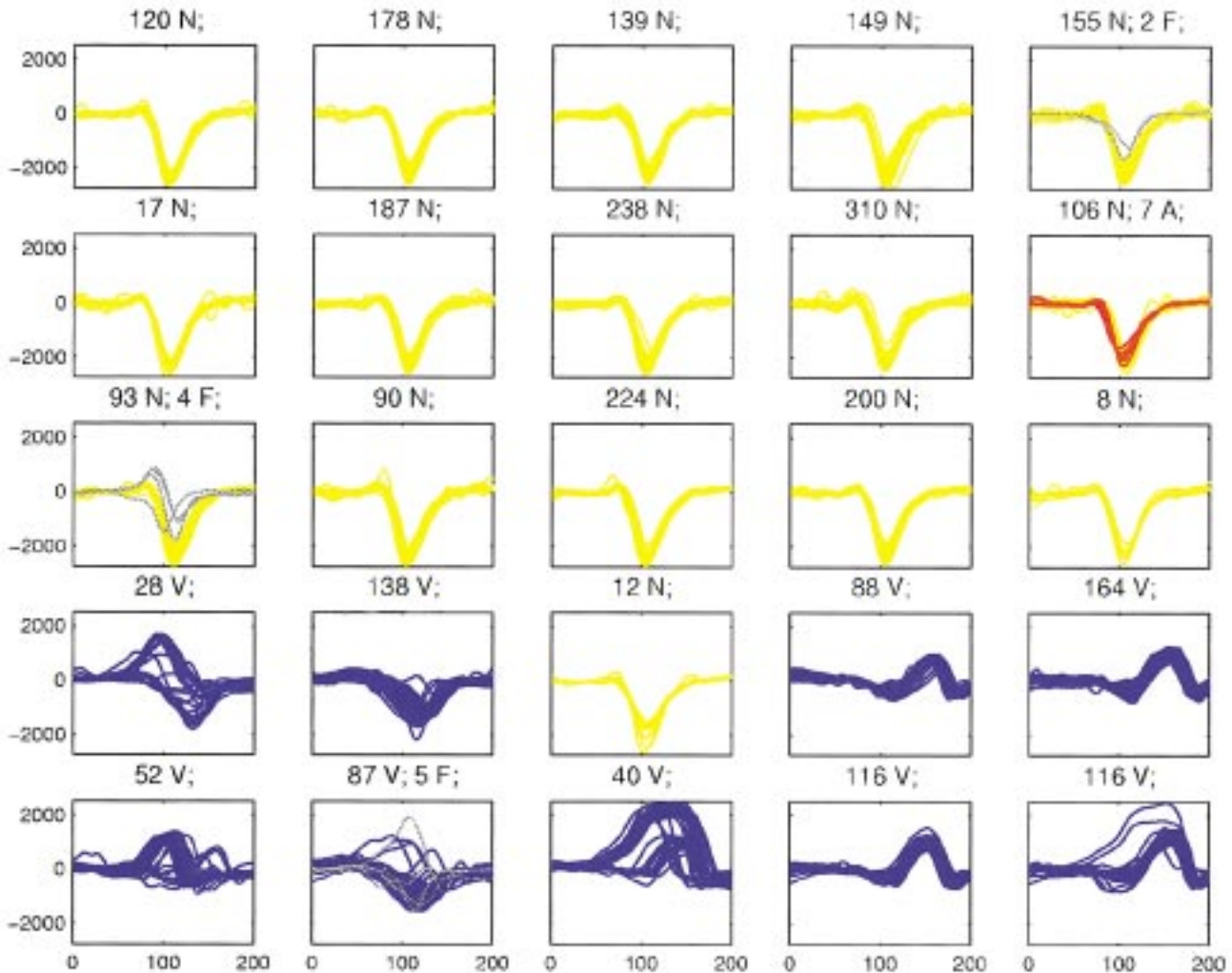


Fig. 5. Beats from a typical 30-min record (#233 lead B) assigned to the different clusters (output nodes). For the notations N, A, V, and F see Table I. The colors denote the different beats according to the expert annotation; N (yellow), A (red), V (blue), and F (black).

With a large neighborhood the cluster centers tend to stick to dense regions, whereas with a lower one it is easier for a cluster centers to leave the neighbors and go far from the dense regions. The choice of T_0 is, thus, a tradeoff between having a well-structured output map and the risk of mixing outliers with the relevant beats.

In phase I, T should not be too low. Hence, T is chosen as

$$T = \max(T_{t-1}c, T_{\min}) \quad (19)$$

where $T_{\min} = 2/\ln(2)$. In phase II there is no lower bound for T .

c) Learning rate: The learning rate η is in phase I annealed (decreased) according to

$$\eta_{t+1} = k\eta_t \quad (20)$$

with $\eta_0 = 1$ and $k = 0.96$. In order for the convergence to be reasonable fast, η is chosen as

$$\eta_t = \max(\eta_{t-1}c, \eta_{\min}) \quad (21)$$

with $\eta_{\min} = 0.01$. In phase II, η is fixed to η_{\min} .

The weights are updated for every pattern. Other updating rates have been pursued with less efficient training as results.

d) Termination criteria: The energy E of the system is defined as

$$E(t) = \sum_{ik} (\xi_k^i - \omega_{i^*k}(t))^2 \quad (22)$$

where i runs over all the input vectors. The fractional change in E from the last block of ten iterations to the previous block defines ΔE . Both phases are terminated when $\Delta E < 0.001$.

V. RESULTS

A. Resulting Maps

Depending upon the random number initialization, different features will settle in different parts of the 5×5 plane. However, the topological relations should be preserved. In Fig. 5, we show the beat assignments of the 5×5 map of output nodes, for one of the 30-min records in the database. Each box represents a cluster of beats. Also indicated is the contents in terms of four kinds of beats; normal beats (N), atrial premature beat (A), ventricular

TABLE V
MISCLASSIFICATIONS IN PERCENTAGE AVERAGED OVER ALL THE AVAILABLE
BEATS FOR DIFFERENT NUMBER OF HERMITE FUNCTIONS

	4 Hermite fcn	5 Hermite fcn	6 Hermite fcn
SOM	1.55 %	1.51 %	1.54 %

premature beats (V), and fusion of ventricular and normal beats (F). In this case, all seven A and 11 F beats end up in clusters where their kind are not dominant. Note that the beats plotted in the boxes are the original beats of lead B, whereas the input to the SOM are composed of the Hermite representation and the R-measures of both leads.

B. Evaluation of Results

The results above indeed look very good. With only a few exceptions the different types of beats ends up in separate boxes. We next turn to a more quantitative evaluation of the results. First, we do this without benchmarking against other unsupervised approaches. Then, we compare with the results of [19]. In Section VI, we will make comparisons with a cross-correlation clustering method.

A beat is defined as misclassified if it ends up in a cluster where the dominant beat is a different one. The resulting total percentage of misclassified beats for all ECG recordings is equal to 1.5%. This number should be compared with a worst total misclassification percentage of 12%, which would be obtained if all beats of each record were classified in a single box.

Results for four, five, and six Hermite functions decomposition are shown in Table V. Table VI shows the result in terms of a “confusion matrix” divided into the different beat types. The majority of beats are normal (68.3%) and, consequently, the likelihood that normal beats will dominate the clusters is high. Most of the normal beats (99.8%) are assigned to clusters in which they represent the dominant fraction. The most common type of misclassification (45.7% of the misclassified beats) is that nonnormal beats are assigned to clusters dominated by the normal ones. Another common type of misclassification is between classes that by definition are very close to each other. The f beats are a mixture between P (paced) and N (normal) beats and the F beats are a mixture between V (ventricular) and N beats. P beats classified as f or vice versa and V beats classified as F or vice versa account for 21.4% of all misclassifications. The unclassifiable beats (Q beats) constitute a very heterogeneous group and, not surprisingly, most of these beats are misclassified.

C. Comparison with a Mixture-of-Expert Model

As mentioned in Section I, in [19] the MIT-BIH arrhythmia database [20] was used for classifying beats using ANN methods. The approach differs from ours both with respect to the overall objectives and the methods employed. The QRS complexes are not automatically identified—rather the database annotations are used. Also, initial labeling of the beats is required. This is contrast to our approach, which handles all the three steps, QRS detection, compact representation and

classification, in an integrated manner. The representation of each beat is also different. Whereas the nine largest principal components computed from 14 samples (180 Hz) are used on each side of the R-peak, we decompose the beats into Hermite functions. Another difference is that we use all 48 records in contrast to [19], where only 33 records (20 for the evaluation), in which premature ventricular contraction can be identified, are kept. Also, the authors of [19] limit their study to two compressed classes, V and non-V, respectively. For the classification, a MOE variant of learning vector quantization is used, whereas in this work we are consistently employing unsupervised learning (clustering). Despite these differences we find it meaningful to make a comparison for the classification part by processing the same records on a 5×5 map.

In order to make comparisons, we need to define a proper error measure. In [19], a performance measure is defined as

$$R_{ME} = \frac{\# \text{ of true negative and true positive beats}}{\text{total } \# \text{ of beats}}. \quad (23)$$

For comparison, we define a similar measure, noting that V beats in [19] correspond to our V and E

$$R^{SOM} = \frac{\text{number of correctly clustered V and E beats}}{\text{total } \# \text{ of beats}}. \quad (24)$$

The comparison between our SOM and the results from MOE [19] are found in Table VII. If an expert were to annotate one typical beat for each of the 25 clusters in our output map while agreeing with the database annotation, one could consider the beats in that cluster as classified to that type, i.e., one could change the word “clustered” in (24) to “classified.” As can be seen from Table VII, within this scenario our method outperforms their method. As a reference a column with the result obtained if all beats were classified as the dominating one (worst case for the SOM approach) is included.

D. Comparison with a Crosscorrelation-Based Method

We also gauge our self-organizing approach against a conventional method based on crosscorrelation (CC) and a set of template beats which has been extensively used in our laboratory [26]. This method can, similar to the SOM, be categorized as unsupervised classification. Due to memory constraints when implementing the CC method, each of the 48 records in the database is divided into three 10-min-long records.

The CC method is initiated by using the first beat as a template beat. Each beat is subjected to linear-phase, bandpass filtering in order to remove frequencies which are judged to be less essential for classification (the -3 -dB filter cutoff frequencies were at 1 and 35 Hz), cf. [27]. Subsequently, each beat is compared to the set of template beats by computing the corresponding CC coefficients: the coefficient is computed by shifting each beat in time until the best correlation is found. A new template beat is created whenever the CC coefficient drops below a noise-dependent threshold value. The threshold is set to 0.94 for low to moderate noise levels and is then gradually decreased to 0.75 for high noise levels (the threshold values were determined from a database not used in this study). This

TABLE VI

THE SOM CLUSTERING OF THE DIFFERENT TYPES OF BEATS USING FIVE HERMITE FUNCTIONS FOR THE REPRESENTATION AND A 5×5 OUTPUT MAP. THE RESULT FROM THE INDIVIDUAL FILES ARE ADDED TOGETHER. THE ANNOTATION FILE TYPES ARE DEPICTED ON THE VERTICAL AXIS, WHEREAS THE CLUSTERS DOMINATED BY THE DIFFERENT TYPES ARE SHOWN HORIZONTALLY. FOR EXAMPLE, 31 OF THE N BEATS ARE ASSIGNED TO CLUSTERS WHERE j IS DOMINANT. $\bar{\Sigma}$ DENOTES (%) THE PERCENTAGE OFF THE DIAGONAL. F IS THE PERCENTAGE OF BEATS IN THE ROW/COLUMN OF ALL BEATS IN THE MATRIX, INDICATING THE IMPACT OF $\bar{\Sigma}$ UPON THE TOTAL RESULT

	N	L	R	A	a	J	S	V	F	b	e	j	E	P	f	Q	$\bar{\Sigma}$	F
N	74699	0	2	49	1	12	0	49	8	0	0	31	0	0	3	0	0.21	68.3
L	0	8046	0	6	0	0	0	4	0	4	0	0	0	0	0	0	0.17	7.4
R	1	0	7095	139	0	0	0	0	0	4	0	0	0	0	0	0	1.99	6.6
A	281	6	34	2213	0	0	0	4	0	0	0	0	0	0	0	0	12.8	2.3
a	13	0	0	9	57	0	0	30	0	0	0	0	0	0	0	0	47.7	0.1
J	7	0	29	0	0	47	0	0	0	0	0	0	0	0	0	0	43.4	0.1
S	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0.0
V	186	3	0	5	1	0	0	6811	79	23	0	0	1	1	0	0	4.21	6.5
F	146	1	3	0	0	0	0	49	599	0	0	0	0	0	0	0	24.9	0.7
b	0	0	0	4	0	0	0	7	0	446	0	0	1	0	0	0	2.62	0.4
e	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0.0
j	75	0	6	0	0	0	0	0	0	0	0	147	0	0	0	0	35.7	0.2
E	1	1	22	0	0	0	0	0	0	16	0	0	66	0	0	0	37.7	0.1
P	1	0	0	0	0	0	0	3	0	0	0	0	0	6942	68	0	1.03	6.4
f	15	0	0	0	0	0	0	1	0	0	0	0	0	158	805	0	17.8	0.9
Q	11	2	0	0	0	0	0	2	0	0	0	0	0	5	11	2	93.9	0.0
$\bar{\Sigma}$	1.00	1.61	1.34	8.74	3.39	20.3	-	2.14	12.7	9.53	-	17.4	2.94	2.31	9.24	0		
F	68.8	7.4	6.6	2.2	0.1	0.1	0.0	6.3	0.6	0.4	0.0	0.2	0.1	6.5	0.8	0.0		

type of threshold design ensures that the creation of new beat classes remains within reasonable limits in noisy signals. The noise level was measured as a root-mean-square value of the high-pass filtered samples contained in the RR interval prior to the QRS complex (high-pass filtering with cutoff frequency at 20 Hz was performed in order to avoid P and T waves to increase the noise level). A beat classified as being similar to an existing class is then used to update the template beat by means of recursive averaging, thus, gradually improving the quality of the template beats.

For each 10-min record, we train a SOM with the same number of clusters (denoted by $A \times B$) as produced by the CC method. We compute the total percentage of misclassified beats for both methods (as defined above). The same is done with fixed map sizes of 5×5 . The results from the CC method is 4.4% misclassified, which should be compared with 2.3% for the SOM under similar conditions (see Table VIII).

VI. DISCUSSION AND CONCLUSION

We have devised a procedure for clustering beats into classes which are not predefined. The proposed method is based on Her-

mite-function encoding of QRS complexes. A self-organized feature map (SOM) is employed for clustering the encoded QRS complexes. The original beats are then presented together with the ones clustered together in a “box.” It is found that the QRS complex of the beats in the MIT-BIH arrhythmia database are accurately detected with our procedure (99.7%). Decomposing the beats into five Hermite functions turns out to be sufficient for achieving a good classification performance. The entire approach is successfully evaluated in three different and independent ways using 5×5 output maps.

- The degree of misclassification is very low (1.5%), where a misclassified beat is defined as a beat ending up in clusters, where the dominant kind is another one. Furthermore, when analyzing the resulting topological maps, these tend to cluster the beats such that similar features are adjacent. The CPU consumption needed for the clustering is modest, less than 1 min/record on a desktop computer (Digital Alpha 250).
- When comparing the clustering results with those from supervised learning methods [19] by classifying each of the 25 clusters according the dominant kind of beat, our method outperforms the supervised approach.

TABLE VII
COMPARISON OF THE MIXTURE OF EXPERT MODEL [19] USING THE R -MEASURES DEFINED IN (23) AND (24). THE REFERENCE COLUMN "REF." SHOWS THE EXTREME CASE OF CLASSIFYING ALL BEATS AS THE DOMINATING ONE

record	R^{SOM} (%)	R^{ME} (%)	ref. (%)
200	98.1	81.0	68.0
201	99.9	95.4	89.7
202	99.4	71.7	99.1
203	96.8	87.4	85.1
205	99.9	97.1	97.4
207	96.9	88.3	90.9
208	98.5	90.5	66.3
209	100.0	98.7	100.0
210	98.3	93.4	92.7
213	97.6	91.9	93.2
214	99.9	98.2	88.7
215	99.9	98.4	95.1
219	99.6	97.4	97.0
221	100.0	99.0	83.7
223	99.1	94.4	81.9
228	100.0	99.0	82.3
230	100.0	99.1	100.0
231	99.9	99.8	99.9
233	99.8	97.7	73.0
234	100.0	99.8	99.9
Avg.	99.1	94.0	88.8

TABLE VIII
RESULTS FROM THE 10-MIN RECORDING USED TO COMPARE SOM WITH THE CORRELATION METHOD. THE SOM IS EVALUATED FOR A 5×5 MAP AND FOR THE SIZE PRODUCED BY THE CORRELATION METHOD, $A \times B$

Method	# misclassified
Correlation	4.4%
SOM $A \times B$	2.3%
SOM 5×5	1.3%

- As a reference we also classify 10-min-long records using a method based on template CC. It turns out that our method performs significantly better than the latter.

The MIT/BIH database was selected because it contains a wide variety of QRS complex morphologies as well as different types of noise and artifacts. Another reason for considering this database was its use in other studies and, thus, comparison of results can be performed. Other annotated databases exist but it seems that these suffer from certain drawbacks, e.g., having a too sparse amount of noise and artifacts (the AHA database, [24]) or being collected with a particular clinical feature in mind (the European ST-T database is primarily concerned with myocardial ischemia, [25]).

The primary motivation for using a clustering algorithm that conserves some of the neighborhood (topology) is that it facilitates the interpretation by cardiologist since similar clusters are presented as neighbors in the map. When it comes to the "hard" results presented, a simpler clustering algorithm without this feature would do as well as the SOM. The benefits for the cardiologist using SOM remains to be exploited. One should mention that our approach not only classifies well. It also embraces all processing steps from the raw ECG signal to the clustered beats; no corrections based on annotations were inserted in between the different steps.

A limitation with the present method is that no information on signal quality was included in the SOM input vector, e.g., as expressed by a signal-to-noise ratio. It is reasonable to assume that the performance will become even better when such measurements are included. Another step which remains to be considered is the classification of beat episodes. The present classification of single beats could be one part of such an analysis but the time relation between the beats and the heart rate also needs to be considered. Also, it could be beneficial to include the encoding error e.g., as a separate input parameter. The SOM could then sort the beats with a high error signal into a separate box.

It is not within the scope of this project to adjust the method to perform particularly well for a certain type of beats. The best performance is found for the normal beats, a class that also outnumbered the other beats. However, it is possible to fine tune the algorithm to give a higher performance for a special type of beat. For example, for ventricular beats one finds a strong correlation to the width of the QRS. One could then utilize this fact by enhancing the importance of width measures by e.g., simply multiplying the corresponding input feature by a factor larger than one.

By increasing the size of the output map one will have more cluster centers and, thus, allow for better separation between beats with different shapes. On the other hand a very large output map will be more difficult to use for the physician. Our believe is that a 5×5 map would be a good tradeoff; the physician could view the whole matrix at once and decide on which beats to take a closer look. The optimal size of the map have to be decided in a future clinical evaluation.

ACKNOWLEDGMENT

The authors would like to thank O.Rauer and R. Rittner for contributions at an early stage of this project and three anonymous referees for valuable suggestions on the manuscript.

REFERENCES

- [1] J. L. Willems, C. Abreu-Lima, P. Arnaud, L. H. van Bemmel, C. Brohet, R. Degani, B. Denis, J. Gehring, I. Graham, G. van Herpen, H. Machado, P. W. Macfarlane, J. Michaelis, S. D. Mouloupos, P. Rubel, and C. Zywiets, "The diagnostic performance of computer programs for the interpretation of electrocardiograms," *New Eng. J. Med.*, vol. 325, pp. 1767–1773, 1991.
- [2] W. G. Baxt, "Application of artificial neural networks to clinical medicine," *Lancet*, pp. 1135–1138, 1995.
- [3] B. Hedén, H. Öhlin, R. Rittner, and L. Edenbrandt, "Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks," *Circulation*, vol. 96, pp. 1798–1802, 1997.
- [4] T. F. Yang, B. Devine, and P. W. Macfarlane, "Use of artificial neural networks within deterministic logic for the computer ECG diagnosis of inferior myocardial infarction," *J. Electrocardiol.*, vol. 27 suppl., pp. 188–193, 1994.
- [5] P. Bozzola, G. Bortolan, C. Combi, F. Pinciroli, and C. Brohet, "A hybrid neuro-fuzzy system for ECG classification of myocardial infarction," in *Computers in Cardiology*, 1996, pp. 241–244.
- [6] H.-S. Chow, G. B. Moody, and R. G. Mark, "Detection of ventricular ectopic beats using neural networks," in *Proc. Computers in Cardiology*, 1992, pp. 659–662.
- [7] M. Morabito, A. Macerata, A. Taddei, and C. Marchesi, "QRS morphological classification using artificial neural networks," in *Proc. Computers in Cardiology*, 1991, pp. 181–184.
- [8] R. Watrous and G. Towell, "A patient-adaptive neural network patient monitoring algorithm," in *Proc. Computers in Cardiology*, 1995, pp. 229–232.
- [9] K. M. Strand, L. R. Smith, M. E. Turner, and J. A. Mantle, "A comparison of simple and template variable models for discrimination between normal and PVC waveforms," *Proc. Computers in Cardiology, IEEE Computer Society*, pp. 21–26, 1980.
- [10] S. H. Rappaport, L. Gillick, G. B. Moody, and R. G. Mark, "QRS morphology classification: Quantitative evaluation of different strategies," in *Proc. Computers in Cardiology*, 1982, pp. 33–38.
- [11] F. Jager, G. B. Moody, S. Divjak, and R. G. Mark, "Assessing the robustness of algorithms for detecting transient ischemic ST segment changes," in *Proc. Computers in Cardiology*, 1994, pp. 229–232.
- [12] T. Y. Young and W. H. Huggins, "On the representation of electrocardiograms," *IEEE Trans. Biomed. Eng.*, vol. BME-10, pp. 86–95, July 1963.
- [13] S. Karlsson, "Representation of ECG records by Karhunen-Loeve expansion," in *Dig. 7th Int. Conf Medical and Biological Engineering*, Stockholm, Sweden, 1967, p. 105.
- [14] N. Ahmed, P. J. Milne, and S. G. Harris, "Electrocardiographic data compression via orthogonal transforms," *IEEE Trans. Biomed. Eng.*, vol. BME-22, pp. 484–487, June 1975.
- [15] L. Sörnmo, P. O. Börjesson, M. E. Nygård, and O. Pahlm, "A method for evaluation of QRS shape features using a mathematical model for the ECG," *IEEE Trans. Biomed. Eng.*, vol. BME-28, pp. 713–717, Nov. 1981.
- [16] P. Laguna, R. Jané, S. Olmos, N. V. Thakor, H. Rix, and P. Caminal, "Adaptive estimation of the QRS complex wave features of the ECG signal by the Hermite model," *Med. Biol. Eng. Comput.*, vol. 34, pp. 58–68, 1996.
- [17] R. Jané, S. Olmos, P. Laguna, and P. Caminal, "Adaptive Hermite models for ECG data compression: Performance and evaluation with automatic wave detection," in *Proc. Computers in Cardiology*, 1993, pp. 389–392.
- [18] T. Kohonen, "Self organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, p. 59, 1982.
- [19] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 891–900, Sept. 1997.
- [20] R. Mark and G. Moody, *MIT-BIH Arrhythmia Data Base Directory*. Cambridge: Massachusetts Inst. Technol., 1988.
- [21] M. E. Nygård, "Delineation of the QRS complex using the envelope of the ECG," *Med. Biol. Eng. Comput.*, vol. 21, pp. 538–547, 1983.
- [22] American Heart Association Committee on Electrocardiography, "Recommendations for standardization and specifications in automated electrocardiography: Bandwidth and digital signal processing," in *Circulation*, 1990, vol. 81, pp. 730–739.
- [23] P. Laguna, R. Jané, P. Caminal, H. Rix, and N. V. Thakor, "Adaptive estimation of the QRS complex wave in the electrocardiographic signal (ECG) by the Hermite model: Classification and ectopic beat detection," *Med. Biol. Eng. Comput.*, vol. 34, pp. 58–68, 1995.
- [24] R. Hermes, D. Geselowitz, and G. Oliver, "Development, distribution and use of the American Heart Association database for ventricular arrhythmia detector evaluation," in *Proc. Computers in Cardiology, III*, 1980, pp. 263–266.
- [25] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi, "The European ST-T Database: Standards for evaluating systems for the analysis of ST-T changes in subsidiary electrocardiography," *Eur. Heart J.*, vol. 13, pp. 1164–1172, 1992.
- [26] L. Sörnmo, R. Atarius, and M. Sunemark, "The ECG signal processing toolbox," Lund Univ. Signal Processing Group, Lund, Sweden, Tech. Rep. SPR-29, 1995.
- [27] J. L. Talmon, "Pattern recognition of the ECG—A structured analysis," master's thesis, Free Univ, Amsteram, The Netherlands, 1983.
- [28] T. Kohonen, *Self-Organization and Associative Memory*, 3rd ed. Berlin, Germany: Springer-Verlag, 1990.



Martin Lagerholm received the Ph.D. degree in theoretical physics and the M.Sc. degrees in physics from Lund University, Lund, Sweden, in 1998 and 1993, respectively.

Since 1998, he works for Foss Tecator, Höganäs, Sweden, on indirect methods (utilizing artificial neural network) for fast quality measurements on agricultural products. His main research interest concerns artificial neural networks and their applications.



Carsten Peterson received the M.Sc. and Ph.D. degrees in theoretical physics from Lund University, Lund, Sweden, in 1972 and 1977, respectively.

He is Professor of Theoretical Physics, Lund University and heads the Complex System Division. His research interests were initially focused on theoretical particle physics and statistical mechanics. Since 1986, his interests has been focused on artificial neural networks, data mining, and time series analysis with some emphasis on biomedical applications, resource allocation problems, Monte Carlo sampling methods, and mean field approximations, thermodynamics of macromolecules, protein folding/design, bioinformatics, and genetic expression networks.



Guido Braccini was born in Florence, Italy, in 1973. He received the M.Sc. degree in electronic engineering in 1997 from the University of Florence, Florence, Italy. He is now working toward the Ph.D. degree in computer science and is working on intelligent agents and electronic commerce.

In 1997, he was a Researcher in the Clinical Physiology Department, Lund University, Lund, Sweden, working on self-organizing maps.



Lars Edenbrandt was born in 1957. He received the M.D. and Ph.D. degrees from Lund University, Lund, Sweden, in 1983 and 1989, respectively. He received Specialist accreditation in clinical physiology 1990 and nuclear medicine 1997, respectively.

He became Associate Professor (Hon), Lund University in 1992. He has been a Consultant with the Department of Clinical Physiology, University Hospital, Lund, Sweden, since 1992. In his research work he has applied artificial neural networks in the field of automated interpretation of electrocardiograms and diagnostic images of the heart.



Leif Sörnmo (S'77–M'84) received the M.Sc. and D.Sc. degrees in electrical engineering from Lund University, Lund, Sweden, in 1978 and 1984, respectively.

He held a part time position with the Department of Clinical Physiology, Lund University, from 1983 to 1995 working with computer-based ECG analysis. He has been an Associate Professor with the Signal Processing Group, Department of Applied Electronics, Lund University, since 1990. His research interests include statistical signal processing and its application to the modeling and analysis of biomedical signals. His recent projects include detection of micropotentials in the ECG, spatiotemporal methods for detection of myocardial ischemia, and time-frequency analysis of atrial fibrillation.