

# A Patient-Adapting Heartbeat Classifier Using ECG Morphology and Heartbeat Interval Features

Philip de Chazal\*, *Member, IEEE*, and Richard B. Reilly, *Senior Member, IEEE*

**Abstract**—An adaptive system for the automatic processing of the electrocardiogram (ECG) for the classification of heartbeats into one of the five beat classes recommended by ANSI/AAMI EC57:1998 standard is presented. The heartbeat classification system processes an incoming recording with a global-classifier to produce the first set of beat annotations. An expert then validates and if necessary corrects a fraction of the beats of the recording. The system then adapts by first training a local-classifier using the newly annotated beats and combines this with the global-classifier to produce an adapted classification system. The adapted system is then used to update beat annotations. The results of this study show that the performance of a patient adaptable classifier increases with the amount of training of the system on the local record. Crucially, the performance of the system can be significantly boosted with a small amount of adaptation even when all beats used for adaptation are from a single class. This study illustrates the ability to provide highly beneficial automatic arrhythmia monitoring and is an improvement on previously reported results for automated heartbeat classification systems.

**Index Terms**—Adaptive classifier, ECG, heartbeat classifier, linear discriminant analysis, statistical classifier model.

## I. INTRODUCTION

THE study of the electrocardiogram (ECG) signal provides an insight to understand decreased heart function for many cardiac conditions. One area of study focuses on arrhythmias, which are any disturbance in the rate, regularity, site of origin or conduction of the cardiac electric impulse [1]. The manifestation in the ECG of life-threatening arrhythmias such as ventricular fibrillation and tachycardia are well understood and a number of successful detection systems have been published (e.g., [2]–[5]).

Other arrhythmias are not imminently life-threatening but may require therapy to prevent further problems. Some arrhythmias appear infrequently, and very long ECG recordings are needed to capture them. Analysis of such a large number of data, usually collected by Holter devices, is very time consuming and, thus, automatic detection and quantification of these arrhythmias is of great assistance to the clinician. In critical care or operating room patient monitoring, computer-assisted real-time classification of arrhythmia is of considerable importance. Many arrhythmias manifest as sequences of heartbeats with

unusual timing or ECG morphology. An important step towards identifying an arrhythmia is the classification of heartbeats. The rhythm of the ECG signal can then be determined by knowing the classification of consecutive heartbeats in the signal [6]. Classification of heartbeats can be very time-consuming and hence any automated processing of the ECG that performs this process would be of assistance and is the focus of this study.

Automated classification of heartbeats has been previously reported by other investigators (e.g., [7]–[14]) using a variety of features to represent the ECG and a number of classification methods. Features include ECG morphology [8], [9], [13], [14], heartbeat interval features [8], [9], [11], [13], [14], frequency based features [7], higher order cumulant features [10], Karhunen-Loeve expansion of ECG morphology [11], and hermite polynomials [12]. Classifiers methods employed include linear discriminants [7], [13], back propagation neural networks [8]–[10], self organizing maps with learning vector quantisation [11], self organizing networks [12] and  $k$ th nearest-neighbours rules [14].

Reporting performance results of cardiac rhythm algorithms has been standardized by the Association for the Advancement of Medical Instrumentation (AAMI) [15], [16]. The AAMI standards emphasise the problems of classifying ventricular ectopic beats (VEB) from non-VEBs and classifying supraventricular ectopic beats (SVEB) from non-SVEBs. However, despite the existence of this standard very few studies [11]–[13] have reported performance results in this format. These three studies all used data from the MIT-BIH arrhythmia database [17]. For this study we have adopted the AAMI standards and utilized the MIT-BIH arrhythmia database in order to compare our results directly to these studies.

Researchers have strived to design heartbeat classification systems that are able to process recordings without any requirement for input from an expert—the so-called global-classifier approach. Such classifiers are generally trained on large databases and endeavour to have good generalisation performance. Unfortunately in practice these systems have not performed well enough and have not been widely used clinically. The reason for underperformance seems to lie in the observation of large inter-patient variation of the morphologies of ECG waveforms.

A selection of results using the global-classifier approach on the MIT-BIH arrhythmia database follows. Hu *et al.* [11] achieved a 62.2% accuracy in separating VEBs from non-VEBs, and we [13] achieved a 97.4% accuracy for the same task. Lagerholm *et al.* [12] described a method for clustering ECG heartbeats from a recording into 25 clusters and determined that on average 98.5% of the heartbeats in any one cluster were from the same heartbeat class. This translates to a classification performance of 98.5% if an expert can correctly

Manuscript received October 3, 2005; revised May 21, 2006. This work was supported in part by the Informatics Research Initiative of Enterprise Ireland. Asterisk indicate corresponding author.

\*P. de Chazal is with the BiancaMed Ltd., NovaUCD, Belfield Innovation Park, University College Dublin, Belfield Dublin 4, Ireland (e-mail: philip.dechazal@biancamed.com).

R. B. Reilly is with the School of Electrical, Electronic and Mechanical Engineering, University College Dublin, Belfield Dublin 4, Ireland (e-mail: richard.reilly@ucd.ie).

Digital Object Identifier 10.1109/TBME.2006.883802

identify the dominant beat of a cluster. Christov *et al.* [14] achieved a 78.3% specificity and 81.6% sensitivity in separating beats originating from the SA node from premature ventricular contractions (PVCs).

In order to improve performance, research attention has been directed to patient adaptive arrhythmia detection, i.e., the classifier uses expert knowledge about a section of the recording under analysis to improve the detection rate on the rest of the recording. Hu *et al.* customized a local heartbeat classifier to a specific patient and then combined it with a global-classifier using a mixture of experts (MOE) approach [11]. The local-classifier requires a cardiologist to annotate a segment of a patient-specific ECG in order to implement the MOE methodology. This MOE approach achieved an accuracy of 94.0% for distinguishing VEBs from non-VEBs. Christov *et al.* [14] also considered a local-classifier approach and achieved 97.7% specificity and 97.3% sensitivity in separating beats originating from the SA node from PVCs. The performance of both Hu's and Christov's locally adapted classifiers were higher than the performance of their global-classifier systems.

## II. AIM

The aim of this paper is to use the global-classifier approach we previously published [13] and combine it with a local-classifier to improve classifier performance. A desired characteristic of the system is that the local-classifier learns as rapidly as possible in order to minimise the required expert cardiologist intervention.

The organization of the paper is as follows. Section III gives a description of the dataset used for training and testing. Section IV provides a detailed description of the methodology in the design of the patient adapting heartbeat classification system. Section V provides the results of tests on various patient adapting classifier designs and the impact of these test results and the conclusions drawn are discussed in Section VI. The overall conclusion of the study is summarized in Section VII.

## III. ECG DATA

The MIT-BIH arrhythmia database [17] was used as the data source in this study and is described in [13]. Briefly, the database contains 48 recordings each containing two ECG lead signals that are band-pass filtered at 0.1–100 Hz and sampled at 360 Hz. There are over 109 000 labeled ventricular beats from 15 different heartbeat types. The heartbeat types were combined into five heartbeat classes using the AAMI recommended practice and these classes were used in all subsequent processing (see [13, Table I]). Class N contains beats originating in the sinus node (normal and bundle branch block beat types), class S contains SVEBs, class V contain VEBs, class F contains beats that result from fusing normal and VEBs, and class Q contains unknown beats including paced beats.

The database also contains a series of manually verified QRS detection points that were utilized in this study.

Using the same methodology as [13], the four recordings containing paced beats were removed from our analysis and the remaining recordings were divided into two equal-sized datasets containing ECG data from 22 recordings (see [13, Table II]). The first dataset (DS1) was used to evaluate the performance to set the values of parameters to optimise performance of the classifier. The second dataset (DS2) was used for an independent

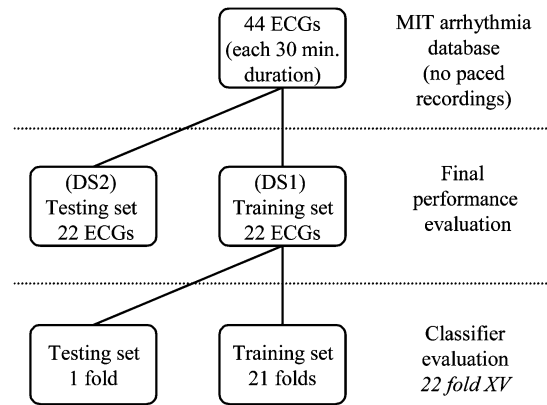


Fig. 1. The division of the MIT-BIH arrhythmia database into training and testing sets to evaluate classifier performance. The 44 nonpaced recordings were divided into two equal size sets—DS1 and DS2. DS1 was used to optimise the classifier configuration using the cross validation (XV) scheme. Once the classifier configuration was optimized, DS2 was used to provide a final independent performance evaluation.

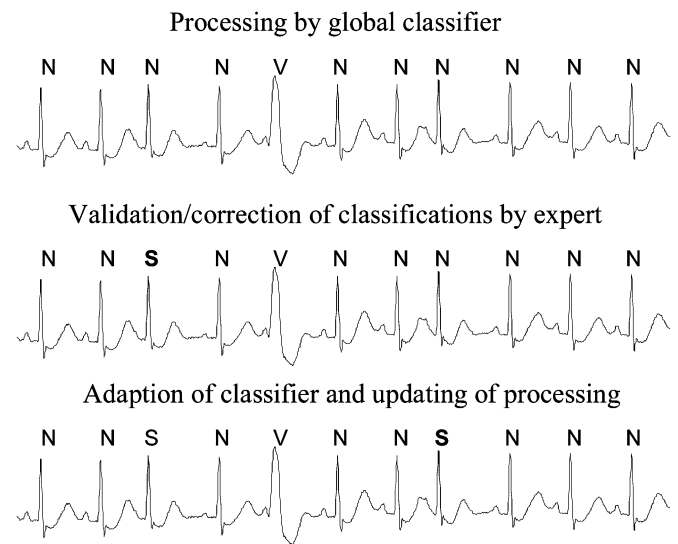


Fig. 2. The adaptation stages of the system. Initially, the recording is processed by the global-classifier to produce the first set of beat annotations. An expert then validates and if necessary corrects the first few beats of the recording (in the example shown the first four beats are checked by the expert and the beat-class of the third beat is changed from N to S). The local-classifier is then trained using the expert annotated beats and combined with the global-classifier and the system is adapted. The adapted system is then used to update beat annotations (in this example the fourth last beat has been changed from N to S as a result of the adaption).

and unbiased performance evaluation of the heartbeat classification system (see Fig. 1).

DS1 contained 51020 heartbeats and DS2 contained 49711 heartbeats.

## IV. METHODS FOR AUTOMATED HEARTBEAT CLASSIFICATION

The heartbeat classification system is a patient adapting system. Initially an incoming recording is processed by the global-classifier to produce the first set of beat annotations (as shown in Fig. 2). An expert then validates and if necessary corrects a fraction of the beats of the recording. The system then adapts by first training a local-classifier using the newly annotated beats and combines this with the global-classifier to

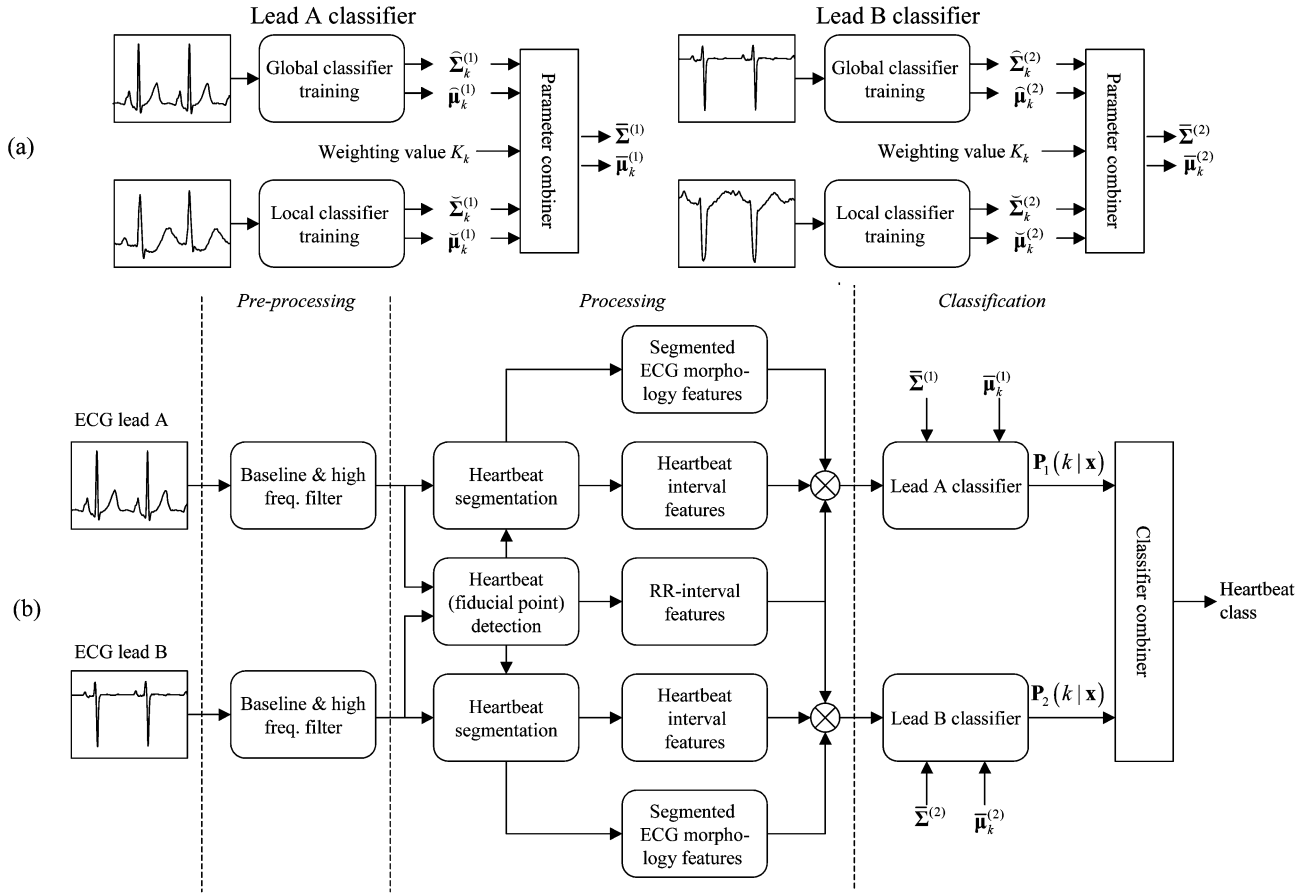


Fig. 3. Schematic representation of an adaptive system for the automated processing of ECG for classification of heartbeats: (a) shows the classifier parameter combining stage and (b) shows the classification stage.

produce an adapted classification system. The adapted system is then used to update the annotations of the beats that have not been annotated by the expert.

#### A. Overview

Fig. 3 depicts the stages of an adaptable automated system suitable for heartbeat classification which is based on the outcomes of this study. It consists of three stages: a pre-processing stage, a processing stage, and a classification stage. The system processes two simultaneously recorded ECG signals and assigns each heartbeat to one of the five AAMI beat classes. The system adaptively updates the parameters of the classifiers as knowledge of the beat classes of the selected beats of the recording under study is included. The pre-processing and processing stages are identical to configuration IX described in [13] and the details are summarized here.

A digital ECG signal is applied at the input to the pre-processing stage. The pre-processing stage utilises a filtering unit to remove artifact signal from the ECG signal. These signals include baseline wander, power line interference and high frequency noise. The processing stage consists of heartbeat detection and feature extraction phases. The heartbeat detection module attempts to locate all heartbeats and is common to the processing of both ECG signals. In this study, the manually verified heartbeat fiducial point times provided with the MIT-BIH arrhythmia database were utilized. Heartbeat segmentation follows heartbeat detection to provide the QRS onset and offset

and T-wave offset times; a Boolean value indicating the presence/absence of a P-wave and, if present, the P-wave onset and offset time for each heartbeat fiducial point.

The feature extraction phase is concerned with forming vectors of measurements (feature vectors) that are processed by the classification stage. A feature vector is calculated from each heartbeat for each ECG signal. The classification stage contains classifier-units that calculate posterior probabilities of the required classes in response to the input feature vectors. A classifier combiner unites the posterior probabilities from the classifier-units and determines the heartbeat class. The classifier-units contain parameters that are set during the system development to optimise the classification performance. The parameters are determined from a combination of local and global training.

A combiner then unites the decisions of the classifier-units from the two ECG signal to form the final decision of the system. The feature extraction phase and the classification stage are discussed in more detail below.

#### B. Feature Extraction

The feature extraction methods were those adopted for the classifier configuration IX system in [13].

Features relating to fiducial point intervals were calculated for each heartbeat. Features relating to heartbeat intervals and ECG morphology were calculated separately for the two ECG signals for each heartbeat [see “Processing” phase of Fig. 3(b)]. Table I lists the features used in this study.

TABLE I  
FEATURES CONSIDERED IN THIS STUDY. WHERE FEATURES ARE DERIVED SEPARATELY FROM THE TWO ECG LEADS, THE GROUP LABELS FOR LEAD B FEATURES ARE SHOWN IN SQUARE BRACKETS

Group Label	Features
RR intervals	<ul style="list-style-type: none"> <li>• Pre-RR interval</li> <li>• Post-RR interval</li> <li>• Average RR-interval</li> <li>• Local avg. RR-interval</li> </ul>
Heart-beat intervals A [B]	<ul style="list-style-type: none"> <li>• QRS duration (QRS offset -QRS onset) of lead A [B]</li> <li>• T-wave duration (T-wave offset - QRS offset) of lead A [B]</li> <li>• P wave flag for lead A [B]</li> </ul>
Morphology A [B]	<ul style="list-style-type: none"> <li>• ECG morphology (10 samples) between QRS onset and QRS offset of lead A [B]</li> <li>• ECG morphology (9 samples) between QRS offset and T-wave offset of lead A [B]</li> </ul>

1) *RR-Interval Features*: Heartbeat fiducial point intervals (henceforth called RR-intervals) were defined as the interval between successive heartbeat fiducial points (the time interval between successive major local extrema). Four features (Table I: RR-intervals) were extracted from the RR sequence. The pre-RR-interval was the RR-interval between a given heartbeat and the previous heartbeat. The post-RR-interval was the RR-interval between a given heartbeat and the following heartbeat. The average RR-interval was the mean of the valid RR-intervals for a recording and had the same value for all heartbeats in a recording. Finally, the local average RR-interval was determined by averaging the valid RR-intervals of the ten RR-intervals surrounding a heartbeat.

2) *Heartbeat Interval Features*: Three features per ECG lead (Table I: Heartbeat intervals A and B) relating to heartbeat intervals were calculated after heartbeat segmentation. The QRS duration was the time interval between the QRS onset and the QRS offset. The T-wave duration was defined as the time interval between the QRS offset and the T-wave offset. The third feature was a Boolean variable indicating the presence or absence of a P-wave.

3) *ECG Morphology Features*: The morphology group (Table I: Morphology A and B) contained amplitude values of the ECG signal determined by a sampling window between the QRS onset and offset and a sampling window between the QRS offset and the T-wave offset points. Ten features were derived by uniformly sampling the ECG amplitude in the first window and nine features were derived by uniformly sampling the second window. As the ECG signals were already sampled, linear interpolation was used to estimate the intermediate values of the ECG amplitude.

In summary, the feature vector processed by the lead A classifier contained the RR intervals, and the lead A heartbeat intervals and morphology features. Similarly, the feature vector processed by the lead B classifier contained the RR intervals, and the lead B heartbeat intervals and morphology features.

### C. Classifier Model

Classifier models based on linear discriminants were utilized throughout this study. The classifier parameters were

determined with the training data using weighted “plug-in” estimates as described in [13]. In the following description of the equations for classifier parameters, the superscripts (1) and (2) appearing in Fig. 3 have been dropped in order to simplify the notation. For a  $c$  class problem the classifier parameters (class means  $\mu_k$  and common covariance  $\Sigma$ ) can be determined from the training data examples using

$$\mu_k = \frac{\sum_{n=1}^{N_k} \mathbf{x}_{kn}}{N_k} \quad (1)$$

and

$$\Sigma_k = \frac{\sum_{n=1}^{N_k} (\mathbf{x}_{kn} - \mu_k)(\mathbf{x}_{kn} - \mu_k)^T}{N_k}$$

$$\Sigma = \sum_{k=1}^c w_k \Sigma_k \quad (2)$$

where the number of training examples in class  $k$  is  $N_k$ , the feature vector of the  $n$ th training example belonging to class  $k$  is denoted  $\mathbf{x}_{kn}$ ,  $w_k$  specifies the weight of class  $k$ , and  $\Sigma_k$  is the class-conditional covariance matrix. The values of  $w_k$  for the five heartbeat classes were set to the same values as used in [13], i.e., N: 400/1608; S: 400/1608; V: 400/1608; F: 400/1608; Q: 8/1608. Note that although (2) is presented in a different way to (4) of [13] both provide the same value for  $\Sigma$ .

Classifier parameters were calculated separately for the local- and global-classifiers and then the parameters were merged (as shown in Fig. 3(a) to form the final adapted system parameters. In the following, the modifier  $\frown$  indicates data or parameters for the global-classifier,  $\checkmark$  indicates local-classifier data or parameters, and  $\checkmark$  indicates adapted classifier parameters.

1) *Training the Global-Classifier*: The global training data ( $\widehat{\mathbf{x}}_{kn}$ ) was used to determine the  $\widehat{\mu}_k$ 's and  $\widehat{\Sigma}_k$ 's for the global-classifier using

$$\widehat{\mu}_k = \frac{\sum_{n=1}^{\widehat{N}_k} \widehat{\mathbf{x}}_{kn}}{\widehat{N}_k} \quad (3)$$

$$\widehat{\Sigma}_k = \frac{\sum_{n=1}^{\widehat{N}_k} (\widehat{\mathbf{x}}_{kn} - \widehat{\mu}_k)(\widehat{\mathbf{x}}_{kn} - \widehat{\mu}_k)^T}{\widehat{N}_k} \quad (4)$$

where  $\widehat{N}_k$  is the number of global training examples in class  $k$ .

During development of the system global training is performed once on a large database and the global-classifier parameters are then fixed.

2) *Training the Local-Classifier*: The local data ( $\checkmark\mathbf{x}_{kn}$ ) was used to determine the  $\checkmark\mu_k$ 's and  $\checkmark\Sigma_k$ 's for the local-classifier using

$$\checkmark\mu_k = \frac{\sum_{n=1}^{\checkmark N_k} \checkmark\mathbf{x}_{kn}}{\checkmark N_k} \quad (5)$$

$$\checkmark\Sigma_k = \frac{\sum_{n=1}^{\checkmark N_k} (\checkmark\mathbf{x}_{kn} - \checkmark\mu_k)(\checkmark\mathbf{x}_{kn} - \checkmark\mu_k)^T}{\checkmark N_k} \quad (6)$$

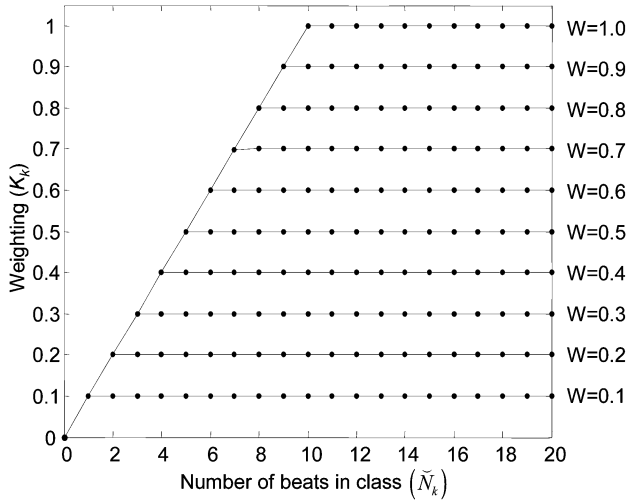


Fig. 4. The relationship between the weighting value  $k(K_k)$  and the number of beats per class to train the local-classifier ( $\check{N}_k$ ) for varying values of  $W$ . A lower weight is given to the local-classifier when less than ten beats per class are used for its training.

where  $\check{N}_k$  is the number of local training examples in class  $k$ . In the event of  $\check{N}_k = 0$ ,  $\check{\boldsymbol{\mu}}_k$  is set to a zero vector and  $\check{\boldsymbol{\Sigma}}_k$  is set to a zero matrix but as explained in the next section this does not influence the classifier training.

The total number of beats used in training of the local-classifier is given by

$$\check{N} = \sum_{k=1}^c \check{N}_k. \quad (7)$$

For each recording under test, the local-classifier parameters are calculated after the expert annotates a section of the recording. In this study, we varied  $\check{N}$  to investigate the effect on performance of the system of different quantities of local-classifier data.

3) *Combining the Local- and Global-Classifier Parameters:* The parameters of the combined classifier were an amalgamation of the local and global parameters. The parameters were combined as follows:

$$\bar{\boldsymbol{\mu}}_k = K_k \check{\boldsymbol{\mu}}_k + (1 - K_k) \widehat{\boldsymbol{\mu}}_k \quad (8)$$

$$\bar{\boldsymbol{\Sigma}}_k = K_k \check{\boldsymbol{\Sigma}}_k + (1 - K_k) \widehat{\boldsymbol{\Sigma}}_k$$

$$\bar{\boldsymbol{\Sigma}} = \sum_{k=1}^c w_k \bar{\boldsymbol{\Sigma}}_k \quad (9)$$

where  $K_k$  is the class conditional weighting value and varies between 0 and 1. As we wished to prevent the adaptive system being too dependent on a small number of local beats in a class, a function was used to calculate  $K_k$ . The function gives a lower weight to the local-classifier when a limited number of beats in a class  $k$  are available (see Fig. 4).

To calculate  $K_k$  we used

$$K_k = \min \left( \frac{\check{N}_k}{10}, W \right) \quad (10)$$

where  $W$  is the ‘‘longterm’’ weighting, i.e., the weighting to be applied when sufficient local beats are available for training.

Values for  $W$  were trialed ( $W = 0.1, 0.2, \dots, 1$ ) and the performance optimizing value selected.

Equations (8)–(10) allow the classifier parameters to adapt even when the data for training the local-classifier does not contain examples of heartbeats from all classes. In the event of no local-classifier data being available for class  $k$  (i.e.,  $\check{N}_k = 0$ ), then  $K_k = 0$  and  $\bar{\boldsymbol{\mu}}_k$  and  $\bar{\boldsymbol{\Sigma}}_k$  simply take the values of  $\widehat{\boldsymbol{\mu}}_k$  and  $\widehat{\boldsymbol{\Sigma}}_k$  calculated from the global training data. Consequently, the system can adapt using just a few beats for the training of the local-classifier and thus minimizing the required input from the expert.

#### D. Classifying and Combining Classifiers

After determining the  $\bar{\boldsymbol{\mu}}_k$ 's and  $\bar{\boldsymbol{\Sigma}}_k$  from the training data, a feature vector  $\mathbf{x}$  is classified by assuming values for the prior probabilities  $\pi_k$  and calculating the estimated posterior probabilities,  $P(k | \mathbf{x})$  for the  $k$ th class using

$$P(k | \mathbf{x}) = \frac{\exp(y_k)}{\sum_{l=1}^c \exp(y_l)} \quad (11)$$

where  $y_k = -(1/2)\bar{\boldsymbol{\mu}}_k^T \bar{\boldsymbol{\Sigma}}^{-1} \bar{\boldsymbol{\mu}}_k + \bar{\boldsymbol{\mu}}_k^T \bar{\boldsymbol{\Sigma}}^{-1} \mathbf{x} + \log(\pi_k)$ .

The prior probability of the N, S, V and F classes was set to 10/41. As the number of unknown beats in general is a tiny fraction of all the heartbeats, the prior probability of the Q class was reduced relative to the other classes and was set to 1/41.

In this study, separate features were calculated from the two ECG leads and each processed to obtain separate posterior probability estimates.

To combine the posterior probabilities estimates the unweighted Bayesian product integration scheme [20], [21] was used. With the classifier outputs from the two ECG leads labelled as  $P_1(k | \mathbf{x})$  and  $P_2(k | \mathbf{x})$  then the final posterior probability output  $\bar{P}(k | \mathbf{x})$  was calculated using

$$\bar{P}(k | \mathbf{x}) = \frac{\prod_{m=1}^2 P_m(k | \mathbf{x})}{\sum_{l=1}^c \prod_{m=1}^2 P_m(l | \mathbf{x})} \quad (12)$$

where  $c$  is the number of classes.

The final classification of the system was obtained by choosing the class with the highest posterior probability estimate. By using information from the two signals more efficient use of the available ECG diagnostic information was made.

#### E. Classification and Clustering Performance Measures

The classification and clustering performance of the classifier configurations was calculated using the same performance measures as described in [13]. Multiway classification accuracy (see [13, Table V(c)]) was used to rank the performance of the classifier during different adaptation conditions. Final assessment of performance of the chosen system was carried out in accordance with the AAMI recommendations and two sets of performance measurements were calculated. These performance measures focus on the ability of algorithms to distinguish VEBs from non-VEBs and SVEBs from non-SVEBs (see [13, Table V(a-b)]). All aggregate performance measures for a dataset were calculated by giving each heartbeat equal weighting (the so called ‘‘gross’’ statistics in [15]). Clustering performance of our system ( $R^{LD}$ ) characterized the ability of

TABLE II  
CLASSIFICATION ACCURACY (%) OF THE NONADAPTED AND ADAPTIVE CLASSIFIERS AS A FUNCTION OF WEIGHTING FACTOR ( $W$ ) AND NUMBER OF BEATS USED TO TRAIN THE LOCAL CLASSIFIER ( $\tilde{N}$ ). THE HIGHEST PERFORMANCE AT EACH  $\tilde{N}$  IS SHOWN IN BOLD. PERFORMANCE WAS DETERMINED USING THE BEATS IN DS1 NOT USED IN TRAINING THE LOCAL CLASSIFIER. THE NUMBER OF BEATS USED FOR PERFORMANCE ASSESSMENT VARIED WITH  $\tilde{N}$  AND RANGED BETWEEN 40020 FOR  $\tilde{N} = 500$  AND 51020 FOR  $\tilde{N} = 0$

$W$	Non-adaptive system		Adaptive system					
	$\tilde{N}$	0	10	20	50	100	200	500
0.1		84.5	90.7	91.1	91.0	91.4	91.3	90.9
0.2		84.5	92.1	92.6	92.5	93.0	93.2	93.2
0.3		84.5	92.3	92.9	93.1	93.6	94.1	94.2
0.4		84.5	<b>92.4</b>	<b>93.1</b>	93.7	94.1	94.8	94.6
0.5		84.5	92.3	93.0	94.0	94.5	95.4	95.1
0.6		84.5	92.1	92.9	<b>94.3</b>	94.6	<b>95.7</b>	95.5
0.7		84.5	91.6	92.4	94.3	<b>94.7</b>	95.7	<b>95.7</b>
0.8		84.5	90.9	91.5	94.0	94.5	95.5	95.6
0.9		84.5	90.1	90.3	93.5	94.2	95.1	95.3
1		84.5	89.2	89.1	92.8	93.6	94.4	95.0

the system to place VEBs and non-VEBs into the correct clusters and was compared to  $R^{\text{SOM}}$  of [12].

#### F. Assessing the Performance of the Classifiers

Classifier performance has been estimated using two methods in this study. In the first method, the cross-validation scheme [22], [23] was applied to DS1. The available data was divided into 22 folds with each fold containing data from one recording. The folds were split into training and testing data with the training data containing the data from 21 folds and the testing data containing the data from the remaining fold. The global-classifier was trained on the training data and the local-classifier was trained on the first  $\tilde{N}$  beats of the test recording. The classifier performance was assessed on the remaining beats of the test recording. The purpose of the first assessment was to determine the value of  $W$  that optimized the classifier performance for each  $\tilde{N}$ .

With the value for  $W$  that optimized system performance determined for each  $\tilde{N}$  an independent performance estimation using DS2 was made for the system. This time the global-classifier was trained using all data in DS1 then each record in DS2 was processed as follows. The first  $\tilde{N}$  beats of a recording were used to train the local-classifier and after combining the local and global-classifier parameters the classifier performance was assessed on the remaining beats of the recording.

Fig. 1 shows how the data was divided in this study into separate sets for performance assessment.

## V. RESULTS

The first step was to use DS1 to determine the optimal  $W$  for each value of  $\tilde{N}$ . Table II show the multiway classification accuracy figures for a nonadaptive system (purely global-classifier) and for an adaptive system using varying  $\tilde{N}$ . For the adapting systems we used the first  $\tilde{N}$  beats to train the local-classifier and

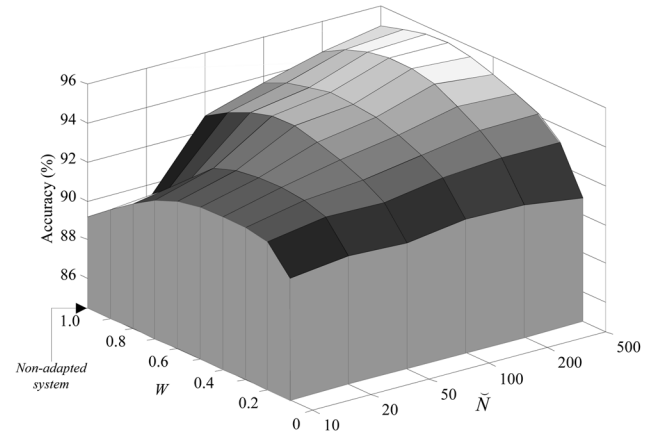


Fig. 5. The classification accuracy on DS1 of the adaptive classifier as a function of weighting factor ( $W$ ) and number of beats used to train the local-classifier ( $\tilde{N}$ ). The  $z$ -axis shows the performance improvement of the adaptive system over the nonadapted system. See the caption for Table II for details on the beats used for accuracy assessment.

TABLE III  
VEB AND SVEB CLASSIFICATION PERFORMANCE ON THE BEATS IN DS2 NOT USED IN TRAINING THE LOCAL-CLASSIFIER FOR VARYING  $\tilde{N}$ . THE NUMBER OF BEATS USED FOR PERFORMANCE ASSESSMENT VARIED WITH  $\tilde{N}$  AND RANGED BETWEEN 38 711 FOR  $\tilde{N} = 500$  AND 49 711 FOR  $\tilde{N} = 0$

	Non-adaptive system		Adaptive system					
	$\tilde{N}$	0	10	20	50	100	200	500
SVEB	Acc	94.6	96.2	96.4	96.2	96.1	96.4	95.9
	Sens	75.9	86.1	88.3	87.4	87.3	87.1	87.7
	+Pred	38.5	49.1	50.8	49.1	48.7	50.6	47.0
	FPR	4.7	3.4	3.3	3.5	3.5	3.3	3.8
VEB	Acc	97.4	98.3	98.5	98.7	98.9	98.9	99.4
	Sens	77.7	81.7	83.6	86.2	87.9	89.2	94.3
	+Pred	81.9	91.1	92.9	93.8	94.4	94.2	96.2
	FPR	1.2	0.6	0.4	0.4	0.4	0.4	0.3

varied  $\tilde{N}$  between 10 and 500. All performance figures were determined by aggregating the test-set figures from the 22 test-set folds. Fig. 5 shows a plot of the same results with the  $z$ -axis height representing the performance improvement of the adaptive system relative to the nonadaptive system.

After consideration of the results in Table II, the best value for  $W$  was selected for each  $\tilde{N}$  and system performance reassessed using DS2 after training the global-classifier on DS1. Table III shows the aggregated classification performance of both the nonadaptive and adaptive systems. Fig. 6 shows a plot of the VEB accuracy for the global and adaptive systems.

Finally for the 500 beat adaptive system, a detailed inspection at the performance in accordance with the AAMI recommendations is provided. The classification performance for each recording and the gross performance figures were calculated for DS2 and are shown in Table IV(a). For comparison the classification performance of a published system [11] is shown in Table IV(b). The clustering performance is shown in Tables IV(c) and V shows a summary table of beat-by-beat performance of the system on DS2.

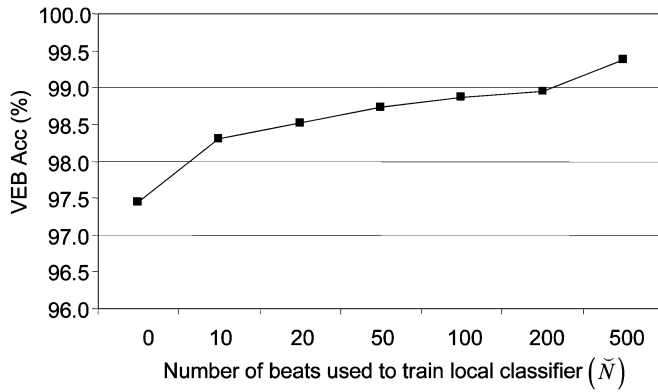


Fig. 6. The VEB accuracy of the nonadapted classifier ( $\tilde{N}$ ) and the adaptive classifier ( $\tilde{N} = 10, 20, \dots, 500$ ) on DS2. See Table III caption for details on beats used for accuracy assessment.

## VI. DISCUSSION

### A. Optimisation of the System on DS1

The system performance was optimized for six adaptive lengths with performance results shown in Table II and Fig. 5. Some interesting trends emerge from these results. First the system performance of the adapted system was always better than the nonadapted system. Remarkably, using just ten beats to train the local-classifier resulted in almost 8% classification accuracy boost in absolute terms. Second, as would be intuitively expected, the more beats that were used to train the local-classifier the better the performance gain. The configuration resulting in the best performance used 500 beats for training the local-classifier and a weighting of 0.7. The multiway classification accuracy of this system was 95.7% which was an 11.2% improvement in absolute terms on the result for the nonadaptive system. There was a diminishing performance boost as more beats were used. Third, as more beats were used to train the local-classifier the optimal weighting factor shifted to favour the local-classifier.

### B. Final Testing of the System on DS2

Table III and Fig. 6 show the classification performance of the optimal systems on the heartbeats in DS2 not used for training the local classifier. This assessment is unbiased as these beats were not used at any point in the development of the classifier. It clearly shows that as the number of beats used to train the local-classifier increases that all the VEB performance figures improve. With no adaptation the system returns a VEB accuracy of 97.4% and this rises to a VEB accuracy of 99.4% when 500 beats are used for adaptation. The same trend for SVEB beats is also seen except that there is small drop off in performance when using 500 beats for the local-classifier.

1) *Detailed System Performance Using 500 Beats for the Local-Classifier:* Table IV(a) shows the performance assessment, as recommended by the AAMI standards, when 500 beats are used to train the local-classifier on DS2. In general, the classifier performed best when there was a variety of beats in the data used for the training the local-classifier.

The summary table of beat-by-beat performance (Table V) provides insight into how groups are being misclassified. These results show that two main errors are being made by the system.

First, 1360 normal (N) beats were misclassified as SVEB (S) beats and second 483 N beats were misclassified as fusion (F) beats. Future work should focus on identifying features that more clearly distinguish normal beats from these two aberrant beat classes. Distinguishing normal beats from fusion beats is an inherently difficult problem as fusion beats are a union of ventricular and normal beats.

### C. Comparison With Other Systems

The VEB classification results of this study were compared to the reported results in Hu *et al.* [11] as this study looked at the problem of distinguishing VEB from non-VEB heartbeats on the MIT-BIH arrhythmia database. Hu used a test-set of twenty recordings which excluded all recordings with no premature ventricular contractions (PVC). There are eleven test-set recordings (see Table IV, footnote “a”) that are common to both studies and the aggregate performance results on these recordings were compared directly. The gross performance of our system using 500 beats to adapt on the same recordings was an accuracy 99.2%, sensitivity 94.4%, positive predictivity 98.4%, and FPR of 0.2%.

Using [11, Table VI] the gross sensitivity, positive predictivity, and FPR rate were recalculated for the 11 recordings in common and results are shown in Table IV(b). Our system is notably better than the fully automatic configuration (GE classifier) approach (accuracy 75.3%, sensitivity 69.6%, positive predictivity 34.6%, and FPR 16.8%) and exceeds the results obtained with their semi-automatic mixture of experts (MOE) approach (accuracy 93.6%, sensitivity 78.9%, positive predictivity 76.0%, and FPR 3.2%). Both systems require approximately the same number of beats to be annotated (our system: 500 beats, Hu’s system—five minutes of recording).

Lagerholm *et al.* [12] also used the same twenty recordings as Hu to determine the performance of clustering VEB and non-VEBs. Using the eleven recordings in common with our study a comparison was made of the clustering performance of our system and Lagerholm’s [see Table IV(c)]. Our system achieved a comparable clustering accuracy ( $R^{LD}$ ) of 99.3% to the 99.2% ( $R^{SOM}$ ) achieved by Lagerholm.

To fully automate the heartbeat classification method presented here an automatic heartbeat detection module is required. Automatic heartbeat detection results in some errors in heartbeat detection (missed heartbeats, erroneously detected heartbeats and errors in heartbeat fiducial point identification) and this will cause a reduction in the performance of the presented heartbeat classifier method.

### D. Utility of the System

The system we have presented is a versatile system. In the absence of an expert, it may be run as a nonadaptive system with a high accuracy of distinguishing VEB, SVEB, and normal beats. Using as few as 10 beats to train a local-classifier and without the need for all beat classes to be represented in the local-classifier training data, the adaptive system performance can be notably improved.

In clinical practice, use of the system could proceed as follows. Initially the nonadaptive system automatically annotates all heartbeats of a presented recording. An expert then reviews a selection of the beat annotations and if satisfied with the annotations then the annotation task is finished. In the event of

TABLE IV  
PERFORMANCE OF THE ADAPTIVE SYSTEM USING 500 BEATS TO TRAIN THE LOCAL-CLASSIFIER. PERFORMANCE IS DETERMINED FOR EACH RECORDING OF DS2 USING THE BEATS NOT USED IN TRAINING THE LOCAL-CLASSIFIER. (a) THE AAMI RECOMMENDED CLASSIFICATION PERFORMANCE MEASURES, (b) AGGREGATE CLASSIFICATION PERFORMANCE FOR SYSTEM IN [11] AND (c) CLUSTERING PERFORMANCE OF OUR SYSTEM AND SYSTEM FROM [12]

															(c)	
Rec	Number of Beats					SVEB				VEB				Clustering		
	N	S	V	F	Q	Acc	Se	+P	FPR	Acc	Se	+P	FPR	$R^{LD}$	$R^{SOM^f}$	
100	1744	28	1	0	0	99.4	67.9	95.0	0.1	100.0	100.0	100.0	0.0	100.0		
103	1582	2	0	0	0	99.9	0.0	-	0.0	100.0	-	-	0.0	100.0		
105	2041	0	26	0	5	100.0	-	-	0.0	98.9	84.6	55.0	0.9	98.9		
111	1623	0	1	0	0	100.0	-	-	0.0	99.9	0.0	-	0.0	99.9		
113	1293	2	0	0	0	100.0	100.0	100.0	0.0	100.0	-	-	0.0	100.0		
117	1034	1	0	0	0	99.9	0.0	-	0.0	100.0	-	-	0.0	100.0		
121	1361	1	1	0	0	99.9	0.0	-	0.0	99.9	0.0	-	0.0	99.9		
123	1016	0	2	0	0	100.0	-	-	0.0	100.0	100.0	100.0	0.0	100.0		
200 <sup>a</sup>	1390	28	681	2	0	98.7	0.0	-	0.0	97.6	93.7	98.9	0.5	97.6	98.1	
202 <sup>a</sup>	1569	54	12	1	0	96.7	0.0	-	0.0	99.5	75.0	60.0	0.4	99.5	99.4	
210 <sup>a</sup>	1959	20	162	9	0	99.1	0.0	-	0.0	97.2	78.4	83.6	1.3	97.2	98.3	
212	2248	0	0	0	0	100.0	-	-	0.0	100.0	-	-	0.0	100.0		
213 <sup>a</sup>	2252	27	199	273	0	99.0	0.0	-	0.0	99.5	94.5	97.9	0.2	99.5	97.6	
214 <sup>a</sup>	1563	0	197	1	0	100.0	-	-	0.0	100.0	100.0	100.0	0.0	100.0	99.9	
219 <sup>a</sup>	1604	3	47	0	0	99.8	0.0	-	0.0	99.6	85.1	100.0	0.0	99.6	99.6	
221 <sup>a</sup>	1626	0	301	0	0	100.0	-	-	0.0	99.9	99.3	100.0	0.0	99.9	100.0	
222	1774	209	0	0	0	89.5	0.0	-	0.0	100.0	-	-	0.0	100.0		
228 <sup>a</sup>	1300	3	250	0	0	99.8	0.0	-	0.0	99.9	100.0	99.6	0.1	99.9	100.0	
231 <sup>a</sup>	1071	0	0	0	0	100.0	-	-	0.0	100.0	-	-	0.0	100.0	99.9	
232	271	1009	0	0	0	99.9	100.0	99.9	0.4	100.0	-	-	0.0	100.0		
233 <sup>a</sup>	1873	4	696	6	0	99.8	75.0	50.0	0.1	99.3	97.6	100.0	0.0	99.3	99.8	
234 <sup>a</sup>	2200	50	3	0	0	98.6	42.0	91.3	0.1	100.0	100.0	100.0	0.0	100.0	100.0	
gross	34394	1441	2579	292	5	95.9	87.7	47.0	3.8	99.4	94.3	96.2	0.3	99.5		
gross <sup>b</sup>										99.2	94.4	98.4	0.2	99.3	99.2	
(b) $GE^b$ [11]										75.3	69.6	34.6	16.8			
$MOE^b$ [11]										93.6	78.9	76.0	3.2			

<sup>a</sup>Recordings in common with the test-set used in [11,12].

<sup>b</sup>Gross statistics calculated using recordings in common with [11,12].

<sup>c</sup> $R^{SOM}$  calculated using equation 24 and Table VII in [12].

TABLE V  
SUMMARY TABLE OF BEAT-BY-BEAT PERFORMANCE FOR THE ADAPTIVE SYSTEM USING 500 BEATS TO TRAIN THE LOCAL-CLASSIFIER FOR DS2. PERFORMANCE DETERMINED USING THE BEATS NOT USED IN TRAINING THE LOCAL-CLASSIFIER

Reference	n	s	Algorithm			
			v	f	q	
N	32432	1360	66	483	53	
S	136	1264	31	10	0	
V	45	65	2433	32	4	
F	24	3	49	216	0	
Q	4	0	1	0	0	

expert not being satisfied, the expert would validate and if necessary override the automatically generated annotations. The local-classifier is then trained on the reviewed section of ECG, the adaptive system updated and the heartbeat annotations of the unreviewed section redone. The expert then reviews another section and repeats the process until satisfied with the annotations.

The average percentage of annotations requiring modification by the expert to achieve adaptation is 100 minus the global multiway accuracy from Table II, i.e.,  $100 - 84.5 = 15.5\%$ . This translates to an average of 1.55 heartbeat annotations requiring change if 10 beats are used to train the local-classifier system

and 77.5 annotations requiring change if 500 beats are used to train the local-classifier system.

## VII. CONCLUSIONS

This paper has discussed the development of a patient adapting heartbeat classifier system. In this system, a local-classifier is trained with a limited number of beat examples from the recording under test and this is then combined with a global-classifier previously trained on a large dataset. With as few as ten beats used for training the local-classifier, the adaptive classifier notably improved on the performance of the nonadaptive classifier. Using more beats to train the local-classifier resulted in greater performance boost although there was a diminishing return. The results of this study illustrate the ability to provide highly beneficial automatic heartbeat classification with minimal expert input.

## REFERENCES

- [1] M. S. Thaler, *The Only EKG Book You'll Ever Need*, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 1999.
- [2] K. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 179-185, Feb. 1999.
- [3] S. Evans, H. Hastings, and M. Bodenheimer, "Differentiation of beats of ventricular and sinus origin using a self-training neural network," *PACE*, vol. 17, pp. 611-626, 1994.



- [4] R. Clayton, A. Murray, and R. Campbell, "Recognition of ventricular fibrillation using neural networks," *Med. Biol. Eng. Comput.*, vol. 32, pp. 217–220, 1994.
- [5] S. Barro, R. Ruiz, D. Cabello, and J. Mira, "Algorithmic sequential decision-making in the frequency domain for life threatening ventricular arrhythmias and imitative artefacts: A diagnostic system," *J. Biomed. Eng.*, vol. 11, pp. 320–328, 1989.
- [6] J. A. Kastor, *Arrhythmias*, 2nd ed. London, U.K.: Saunders, 1994.
- [7] L. Senhadji, G. Carrault, J. J. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 2, pp. 167–173, Mar.–Apr. 1995.
- [8] T. H. Yeap, F. Johnson, and M. Rachniowski, "ECG beat classification by a neural network," in *Proc. Annu. Int. Conf. IEEE Engineering Medicine and Biology Society*, 1990, pp. 1457–1458.
- [9] Y. H. Hu, W. J. Tompkins, J. L. Urrusti, and V. X. Afonso, "Applications of artificial neural networks for ECG signal detection and classification," *J. Electrocardiol.*, vol. 26, pp. 66–73, 1993.
- [10] S. Osowski and T. L. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 11, pp. 1265–1271, Nov. 2001.
- [11] 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, no. 9, pp. 891–900, Sep. 1997.
- [12] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sormmo, "Clustering ECG complexes using hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 838–848, Jul. 2000.
- [13] P. de Chazal, M. O'Dwyer, and R. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.
- [14] I. Christov, I. Jekova, and G. Bortolan, "Premature ventricular contraction classification by the  $k$ th nearest-neighbours rule," *Physiol. Meas.*, vol. 26, pp. 123–130, 2005.
- [15] Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms Association for the Advancement of Medical Instrumentation, Arlington, VA, 1998, ANSI-AAMI EC57:1998 (American National Standard).
- [16] Recommended Practice for Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms Association for the Advancement of Medical Instrumentation, Arlington, VA, 1987, ANSI ECAR:1987.
- [17] R. Mark and G. Moody, MIT-BIH Arrhythmia Database 1997 [Online]. Available: <http://ecg.mit.edu/dbinfo.html>
- [18] P. Laguna, R. Jané, and P. Caminal, "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database," *Comput. Biomed. Res.*, vol. 27, no. 1, pp. 45–60, 1994.
- [19] R. 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 *Proc. Computers in Cardiology*, 1997, vol. 24, pp. 295–298.
- [20] M. Heckmann, F. Berthommier, and K. Kroschel, "Noise adaptive stream weighting in audio-visual speech recognition," *Eurosip J. Appl. Signal Process.*, vol. 11, pp. 1260–1273, 2002.
- [21] I. Bloch, "Information combination operators for data fusion: A comparative review with classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 26, no. 1, pt. A, pp. 52–67, Jan. 1996.

- [22] R. Kohavi, "A study of cross validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artificial Intelligence*, 1995, pp. 1137–1143.
- [23] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford Univ. Press, 1995.



**Philip de Chazal** (M'94) received the B.E. degree in electronic engineering and the M.Biomed.E. and Ph.D. degrees in biomedical engineering from University of New South Wales, Sydney, Australia, in 1989, 1995, and 1999, respectively.

He is cofounder and currently Chief Technical Officer of BiancaMed, Dublin, Ireland, a company providing intelligent computer-based analysis of signals sensed from the human body. Previously, he was a Research Scientist for the CSIRO, Sydney, a Biomedical Engineer for MedCare Systems, Sydney, and a

Research Fellow at University College Dublin, Dublin, Ireland. His research interests include signal processing and pattern recognition for biomedical applications and image processing for multimodal applications.

Dr. de Chazal is a member of the IEEE Engineering in Medicine and Biology Society. He is a reviewer for IEEE TRANSACTIONS ON MULTIMEDIA and IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING.



**Richard B. Reilly** (M'92–SM'04) received the B.E., M.Eng.Sc., and Ph.D. degrees in 1987, 1989 and 1992, all in electronic engineering, from University College Dublin, Dublin, Ireland.

In 1988, he joined Space Technology Ireland and the Department de Recherche Spatiale (CNRS group), Paris, France, developing DSP-based on-board experimentation for the NASA satellite WIND. In 1990, he joined the National Rehabilitation Hospital and in 1992 became a Post-Doctoral Research Fellow at University College Dublin, focusing on signal processing for speech and gesture recognition. Since 1996, he has been on the academic staff in the Department of Electronic and Electrical Engineering at University College, Dublin. He is currently Associate Professor in the School of Electrical, Electronic and Mechanical Engineering at University College Dublin and researches into neurological signal processing and multimodal signal processing. He was the 1999/2001 Silvanus P. Thompson International Lecturer for the Institution of Electrical Engineers (IEE).

In 2004, Dr. Reilly was awarded a US Fulbright Award for research collaboration into multisensory integration with the Nathan Kline Institute for Psychiatric Research, New York. He is a member of the IEEE Engineering in Medicine and Biology Society and Signal Processing Society. He is the Republic of Ireland representative on the Executive Committee of the IEEE United Kingdom and Republic of Ireland Section. He was an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA. He is also a reviewer for IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the *Journal of Applied Signal Processing*, *Signal Processing*, and the *IEE Proceedings Vision, Image & Signal Processing*.