# FEATURE EXTRACTION FOR HEARTBEAT CLASSIFICATION USING INDEPENDENT COMPONENT ANALYSIS AND MATCHING PURSUITS

Germán Gómez Herrero<sup>1</sup>, Atanas Gotchev<sup>1</sup>, Ivaylo Christov<sup>2</sup>, Karen Egiazarian<sup>1</sup>

<sup>1</sup>Tampere University of Technology, Tampere, Finland <sup>2</sup>Centre of Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria

### ABSTRACT

In this paper, we present a method based on the Matching Pursuits algorithm for extraction of time-frequency features that can be used for classification of various abnormal heartbeats. Further, we investigate the usefulness of Independent Component Analysis for extracting additional spatial features from multichannel electrocardiographic recordings. The performance of these two different sets of features is assessed using the 48 recordings of the MIT-BIH arrhythmia database.

## 1. INTRODUCTION

Analysis of the electrocardiogram (ECG) for detecting different types of heartbeats is of major importance in the diagnosis of cardiac dysfunctions. Some arrhythmias appear infrequently, and very long ECG recordings, the so called Holter ECG, are needed to capture them. Analysis of such large amount of data is very time consuming and, therefore, automatic processing systems can be of great assistance to the clinician. In critical care or operating room patient monitoring, computer-assisted real-time classification of abnormal ventricular beats is of considerable importance.

Automatic detection of abnormal cardiac conditions has been previously done using a variety of features to represent the ECG waveform. Most popular approaches are based on pattern recognition techniques using morphological features of the different ECG complexes (e.g. [1,2]). However, several approaches using different sets of features to represent the ECG can be found in the literature (e.g. Hermite functions [3], cumulant features [4], wavelets [5], etc). Despite the high detection accuracies obtained by morphological approaches they have several disadvantages. First, they rely on threshold-based segmentation techniques to separate the different ECG complexes. These techniques are highly sensitive to the large morphological variations of the ECG not only between different patients or patient groups but also within the same patient. Second, the class of waveforms that can be effectively characterized by such kind of features is limited. Some methods use only the morphology of the QRS complex [2], while others include also morphological features of the P and T waves [1]. However, we cannot use morphological features of the P, T or QRS waves to describe cardiac patterns that do not have clear P, T and QRS complexes. Therefore, morphological features are not suitable for describing ventricular fibrillation and some types of tachycardia.

Time-frequency features have been previously used to characterize the ECG [4], [5]. Those studies use linear expansions of the ECG waveforms into a single time-frequency basis such as a wavelet basis. However, it is clear that some ECG patterns are better described in terms of their frequency content (expansion in a Fourier basis) while others have well localized temporal structures and are better characterized by their time representation (expansion in a Dirac basis). In this paper we propose a general method for selecting the time-frequency basis that is best suited to the detection of different beat patterns. The procedure is based on the Matching Pursuit algorithm [6].

Spatial information contained in multichannel ECG recordings can also be used to extract features for heartbeat classification. In this contribution we used Independent Component Analysis (ICA) for the extraction of the spatial features.

## 2. METHODS

Fig. 1 depicts an automated heartbeat classification system based on the time-frequency and ICA features proposed in this study. We used, for training and testing, the MIT-BIH arrhythmia database [7] consisting of two leads ECG recordings. The heartbeats are detected using the fiducial points of the database and extracted using a window of samples of fixed size around each fiducial point. Using the two ECG channels, the ICA block estimates, for each heartbeat, a projection matrix minimizing the statistical dependence among the dimensions of the projected data. The components of this matrix are used as features for the classification stage. The time-frequency feature extraction projects every heartbeat into different sets of wavelet packet atoms that are selected to match the characteristic structures of the different types of beats we are trying to classify. Finally, the time-frequency and ICA features are classified using neural networks.

## 2.1. Matching Pursuit Denoising

Let  $(s_t)_{t=0}^{n-1}$  be a discrete-time 1-dimensional noisy signal and  $\mathcal{D}$  a dictionary of waveforms  $(\phi_{\gamma})_{\gamma \in \Gamma}$ , with  $\gamma$  being the indexing parameter of the dictionary. The waveforms  $\phi_{\gamma}$  are discrete-time 1-dimensional signals of length *n* called atoms. Using columnwise notation, we denote  $\phi_{\gamma} = [\phi_{\gamma,0}, ..., \phi_{\gamma,n-1}]^T$  and  $\mathbf{s} = [s_0, ..., s_{n-1}]^T$ . We face the problem of finding an approximate decomposition of  $\mathbf{s}$ :

$$\mathbf{s} = \sum_{i=1}^{m} \alpha_{\gamma_i} \boldsymbol{\phi}_{\gamma_i} + \mathbf{r}^{(m)} \tag{1}$$

such that  $\hat{\mathbf{s}}^{(m)} = \sum_{i=1}^{m} \alpha_{\gamma_i} \phi_{\gamma_i}$  is the estimated noise-free signal,  $\mathbf{r}^{(m)} = \begin{bmatrix} r_0^{(m)}, ..., r_{n-1}^{(m)} \end{bmatrix}^T$  is a  $n \times 1$  column vector containing the noise, and  $(\alpha_{\gamma_i})_{i=0}^{n-1}$  are a set of scalar coefficients. Matching Pursuit [6] is a greedy algorithm for finding such kind of approximate decompositions. The specific Matching Pursuit denoising scheme



Fig. 1. Scheme of the proposed system for automatic classification of ECG heartbeats

used in this study is described by the following iteration:

$$\begin{split} \hat{\mathbf{s}}^{(0)} &= 0\\ \mathbf{r}^{(0)} &= \mathbf{s}\\ i &= 1\\ \mathbf{while} \ (i \leq m)\\ &\gamma_i = \arg \max_{\gamma} \left\| \left\langle \mathbf{r}^{(i-1)}, \phi_{\gamma} \right\rangle \right\|_1\\ &\alpha_{\gamma_i} &= \left\langle \mathbf{r}^{(i-1)}, \phi_{\gamma_i} \right\rangle\\ &\hat{\mathbf{s}}^{(i)} &= \hat{\mathbf{s}}^{(i-1)} + \alpha_{\gamma_i} \phi_{\gamma_i}\\ &\mathbf{r}^{(i)} &= \mathbf{s} - \hat{\mathbf{s}}^{(i)}\\ &i &= i+1 \end{split}$$
end

where  $\|\cdot\|_1$  denotes the  $l^1$  norm and  $\langle \cdot \rangle$  is the inner product. Therefore, Matching Pursuit selects the m atoms among the available waveforms in the dictionary that best correlate the signal structures. The iteration can be stopped when the coefficient associated with the selected atom falls below a threshold [8] or when a certain number of atoms is reached. We decided to use the latter approach in order to be sure of considering as useful signal, not only the QRS complex, but also the less energetic P and T waves. For determining the number of required atoms to describe a Normal Sinus Beat (NSB) we used the NSBs in the first channel of recording 100 of the MIT-BIH database (which show a high signal to noise ratio). Using only 10 atoms of a Symmlet 8 wavelet packet dictionary, the average residual energy for those beats was as low as 1.45%. In order to have a security margin and guarantee that not any important features of other heartbeat types are discarded during the denoising, we decided to choose 15 atoms as stopping criterion for the Matching Pursuits iteration.

#### 2.2. Time-Frequency feature extraction

For the time-frequency feature extraction we employ wavelet packets because they have high localization power and represent overcomplete dictionaries. However, the same feature extraction procedure can be used with any other type of dictionaries (e.g. Gabor, cosine packets, stationary wavelets, etc). Before proceeding with the actual feature extraction several subdictionaries of wavelet packet atoms matching the characteristics of the different types of heartbeats are constructed. This is done using training data and according to the following steps:

1. The set of heartbeats contained in the training dataset are denoised using the Matching Pursuit algorithm described above. Let  $\hat{s}^{(j)}$  be the denoised *j*th heartbeat and let  $\Gamma$  denote the set of indexes for the full wavelet packet dictionary and  $\Gamma^{(j)}$  the set of indexes corresponding to the atoms that were selected during the Matching Pursuit denoising procedure for the *j*th beat. Then, we can express each denoised heartbeat as:

$$\hat{\mathbf{s}}^{(j)} = \sum_{\gamma \in \Gamma} \alpha_{\gamma}^{(j)} \boldsymbol{\phi}_{\gamma} \tag{2}$$

where  $\alpha_{\gamma}^{(j)} = 0 \ \forall \gamma \notin \Gamma^{(j)}$ .

2. Let  $C_k$  denote the set of indexes of the training heartbeats belonging to class k. We select the m wavelet packet atoms  $(\phi_{\gamma_i})_{i=1}^m$  that best match the structures of class  $C_k$  in the following way:

$$\begin{split} i &= 1 \\ \mathbf{while} \ (i \leq m) \\ \gamma_i &= \arg \max_{\gamma} \sum_{j \in \mathcal{C}_k} \left\| \alpha_{\gamma}^{(j)} \right\|_1 \\ \alpha_{\gamma_i}^{(j)} &= 0 \ \forall j \in \mathcal{C}_k \\ i &= i+1 \\ \mathbf{end} \end{split}$$

The procedure above ranks the atoms according to their average  $l^1$  norm across the beats belonging to class  $C_k$ . The atoms having the greatest average  $l^1$  norm are assumed to represent the most important and stable signal structures.

Once the feature extraction stage has learned the structures of each beat type the actual feature extraction can take place. For N possible heartbeat types, let  $(\Gamma_{C_k})_{k=1}^N$  be the N sets of indexes of the atoms that were selected for representing the characteristic features of the N classes. Then, each heartbeat to be classified is decomposed into these N underdetermined dictionaries of wavelet packets. Let s denote the column vector containing the heartbeat samples and  $\Phi_{\Gamma_{C_k}}$  the matrix whose columns are the atoms selected for class  $C_k$ . Then, this decomposition can be performed by:

$$\mathbf{s} = \Phi_{\Gamma_{\mathcal{C}_{k}}} \alpha_{\mathcal{C}_{k}} + \mathbf{r}_{\Gamma_{\mathcal{C}_{k}}}$$
$$= \Phi_{\Gamma_{\mathcal{C}_{k}}} \left( \Phi_{\Gamma_{\mathcal{C}_{k}}}^{\dagger} \mathbf{s} \right) + \left( \mathbf{s} - \Phi_{\Gamma_{\mathcal{C}_{k}}} \left( \Phi_{\Gamma_{\mathcal{C}_{k}}}^{\dagger} \mathbf{s} \right) \right)$$
(3)

where  $\dagger$  denotes the Moore-Penrose pseudoinverse. We propose to use as features for heartbeat classification the *m* expansion coefficients  $\alpha_{C_k}$  as well as the normalized energy of the residual:

$$\beta_{\mathcal{C}_{k}} = \frac{\left\|\mathbf{r}_{\Gamma_{\mathcal{C}_{k}}}\right\|^{2}}{\left\|\mathbf{s}\right\|^{2}} \tag{4}$$

If a heartbeat belongs to class  $C_k$  then the atoms with indexes in the set  $\Gamma_{C_k}$  are, probably, well correlated with its underlying structures and  $\beta_{C_k}$  should be small. On the contrary, if a beat does not belong to class  $C_k$  then, most likely, its local structures do not correlate well with the atoms with indexes in  $\Gamma_{C_k}$  and  $\beta_{C_k}$  is expected to be large.

#### 2.3. ICA feature extraction

ICA can be defined as a statistical technique for projecting multivariate signals into independent directions. It can be shown [9] that ICA is able to solve the basic Blind Source Separation (BSS) problem in which a set of linear mixtures  $x_1(t), ..., x_m(t)$  of as many unknown independent sources  $s_1(t), ..., s_m(t)$  are observed and we want to determine the source signals given only the mixtures, i.e., given the system:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{5}$$

where  $\mathbf{s} = [s_1(t), ..., s_m(t)]^T$ ,  $\mathbf{x} = [x_1(t), ..., x_m(t)]^T$  and  $\mathbf{A}$  is the unknown mixing matrix, we want to estimate a demixing matrix  $\hat{\mathbf{W}} = \hat{\mathbf{A}}^{-1}$ . In order to solve this estimation problem, ICA assumes the sources to be non-gaussian mutually independent random variables and the mixing matrix to be full rank. Under these assumptions, the mixing matrix  $\mathbf{A}$  can be estimated up to a row permutation and scale factor.

It is generally assumed that the signals recorded in a set of ECG channels can be quite accurately modeled as a linear (instantaneous) mixture of some independent cardiac electrical sources. Different heartbeat types are likely to have different spatial localizations of the electrical sources generating the ECG. These different spatial localizations may result in different electrical transmission matrices and therefore different estimated mixing matrices. Based on this idea, we propose to use the estimated mixing matrix  $\hat{A}$  as features for cardiac beat classification. Computer simulations

using the MIT-BIH arrhythmia database showed that the discrimination performance of the ICA features is especially good for beat types having very different spatial localization of their electrical generators. This is the case e.g. of the paced beats.

## 2.4. Classifier

The classification stage, depicted in Fig. 1, consists of a set of several classifiers trained for detecting a certain class of heartbeats and reject any other type. The input to the classifier trained to detect class  $C_k$  are the expanding coefficients  $\alpha_{C_k}$  and the normalized residual energy  $\beta_{C_k}$  when decomposing the current heartbeat as shown in Eq. 3.

We used as classifiers 3-layers Neural Networks with resilient backpropagation learning algorithm. All the neurons had hyperbolic tangent sigmoid transfer function. The number of neurons in the input layer was equal to the number of input features, there were half as many neurons in the hidden layer and a single neuron in the output layer.

#### 3. RESULTS

#### 3.1. ECG database

We have tested the performance of the proposed features for heartbeat classification using the 48 recordings from the MIT-BIH arrhythmia database [7]. Each recording contains two leads of 30min ECG signals. The data are bandpass filtered at 0.1-100 Hz and the sampling frequency is 360 Hz. 50% of the available heartbeats were used for training and 50% for testing the performance of the trained classification system.

### 3.2. Classification of NSB and PVC

In the first experiment we use only the heartbeats belonging to the classes Normal Sinus Beat (NSB) and Premature Ventricular Contraction (PVC). The performance was assessed in terms of sensitivity (correct detection of NSB) and specificity (correct detection of PVC). Fig. 2 shows the sensitivity and the specificity of the classifier when using only the time-frequency features and varying the number of features used.

The optimum feature extraction stage used 10 time-frequency atoms for each channel and each beat type. The sensitivity obtained for that configuration was 99.9% and the specificity 96.1%. When we included the ICA features in the classification process, the specificity of the system increased to 97.9% and the sensitivity slightly decreased to 99.8%. Therefore, we can conclude that the best performance was obtained with the configuration using 10 time-frequency atoms to describe each class and the mixing matrix coefficients estimated by ICA. This accounted for a total of 48 features, i.e. 10 expansion coefficients and 1 normalized residual energy coefficient for each channel and class, plus the 4 elements of the estimated mixing matrix.

These results are similar to the obtained with the best classifier based on heartbeat interval and QRS morphological features found in [2], where the accuracy of detecting NSBs and PVCs was of 99.7% and 98.5% respectively. Although in [2] only 26 features were used for classification and our system uses 48 features, our approach is more flexible and presents many possibilities for improvement that have not been investigated yet. On the contrary, morphological and heartbeat interval features have already been



Fig. 2. Classification performance of the NSB/PVC classifier versus the number of time-frequency atoms used

class	Number	Sens.	Sens.	Spec.	Spec.
	of beats	(1)	(2)	(1)	(2)
NSB	30139	99.74	99.63	98.79	99.23
LBBB	3206	99.06	99.34	99.96	99.94
RBBB	2895	99.55	99.48	99.94	99.94
PVC	2883	94.97	95.32	99.82	99.85
PB	2848	99.58	99.68	99.93	99.74

**Table 1**. Performance of the multiclass classifier in terms of Sensitivity and Specificity: (1) when using only the time-frequency features and (2) when using the time-frequency and ICA features.

extensively studied in [1] and [2] and their performance has been highly optimized.

#### 3.3. Classification of NSB, LBBB, RBBB, PVC and PB

In the second experiment, we test the classification performance when considering the 5 largest classes of heartbeats in the MIT-BIH arrhythmia database: Normal Sinus Beats (NSB), Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Premature Ventricular Contraction (PVC) and Paced Beats (PB). We used 10 atoms per class and channel and the ICA features. The results compared with the results obtained when we do not include the ICA features are shown in Table 1 . We can observe that the ICA features improve the system's ability of discriminating between NSBs and abnormal beats which can be of considerable importance in clinical use.

The performance of the system using both time-frequency and ICA features is remarkably good, with specificities and sensitivities for the different classes over 99% except for the PVC class. The relatively low specificity for the PVCs can be explained by the fact that, in the case of ventricular pacing, the QRS complex may be wide and bizarre, and the T wave may be inverted, as in the case of PVCs. This makes the separation between ventricular PBs and PVCs complicated.

### 4. CONCLUSIONS

In this paper we have proposed a new method for selecting timefrequency features that can be used for heartbeats classification. We also introduced ICA as a feature extractor for the processing of the ECG. Using the MIT-BIH arrhythmia database and the proposed features, simulation results showed a classification accuracy of 99.8 % for NSBs and 97.9 % for PVCs. Furthermore, the system performance was also very good when classifying 5 types of beats: NSB, LBBB, RBBB, PVC and PB, with classification accuracies over 95% for the 5 beat types. The computing power required by the proposed system is very high during the training of the feature extractors. However, once the system has been trained, the extraction of the time-frequency features only requires the projection of the new beats into the selected wavelet packets atoms. Calculation of the projection coefficients by simple matrix multiplication can be done in a time proportional to O(MLNn), M being the number of data classes, L the number of ECG channels, N the number of atoms used, and n the number of samples of each heartbeat. The ICA feature extraction requires more complex operations but there exist fast ICA algorithms (e.g. JADE [10]) that are suitable for real-time operation when we have to process just few ECG channels.

## 5. REFERENCES

- P. Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1196–1206, 2004.
- [2] I. Christov and G. Bortolan, "Ranking of pattern recognition parameters for premature ventricular contractions classification by neural networks," *Phys. measur.*, vol. 25, pp. 1281–1290, 2004.
- [3] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ECG complexes using hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, pp. 838–848, 2000.
- [4] S. Osowski and T. L. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 1265–1271, 2001.
- [5] L. Senhadji, G. Carrault, J. J. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," *IEEE Eng. Med. Bio. Mag.*, vol. 14, pp. 167–173, 1995.
- [6] S. G. Mallat and Z. Zhang, "Matching pursuits with timefrequency dictionaries," *IEEE Trans. Sig. Proc.*, vol. 41, pp. 3397–3415, 1993.
- [7] R. Mark and G. Moody, "MIT-BIH arrhythmia database," 1997, Available online: http://ecg.mit.edu/dbinfo.html.
- [8] S. Chen, *Basis Pursuit*, Ph.D. thesis, Department of Statistics, Standford University, 1995.
- [9] J.-F. Cardoso, "Blind signal separation: Statistical principles," in *Proceedings of the IEEE*, 1998, vol. 86, pp. 2009– 2025.
- [10] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," in *IEEE Proceedings*, 1993, vol. 140, pp. 362–370.