FUZZY - WEIGHTED AVERAGING FOR HIGH-RESOLUTION ECG BASED ON EXPLORATORY DATA ANALYSIS

N. Laskaris[1], S. Fotopoulos[2], A. Bezerianos[1], A. Manolis[3]
Department of Medical Physics[1] / Physics[2] / Cardiologic clinic[3]
University of Patras, GR-26500, Patras, GREECE
Tel.: +30 61 996115, FAX: +30 61 997745, email: bezer@upatras.gr

ABSTRACT

In this work we introduce a method for the enhancement of Late Potentials in the Signal Averaged electrocardiography. The method involves computation of weights prior averaging. Two fuzzy control techniques are proposed for the derivation of weights. The experimental results indicate the contribution of the method to a more reliable prognosis.

INTRODUCTION

The analysis of Ventricular ***Late Potentials*** (**LP**) is a valuable noninvasive method to identify patients with an increased risk for ventricular tachycardia. **LP** are low-level signals in the terminal portion of the *QRS*, demanding ***High Resolution electrocardiography*** (**HRECG**) for their detection and quantification.

In the **HRECG**, the *QRS* offset is masked by noise. This noise is primarily due to skeletal muscle movement, with smaller contributions from nervous system activity and instrumentation noise.

Signal Averaging is the widely accepted procedure to reduce the noise while preserving the **LP** [1]. The purpose of the ***signal averaged ECG*** (**SAECG**) is to enhance the cardiac *Signal to Noise Ratio* (SNR) so as to optimize the measurements referred to **LP** (e.g time-amplitude analysis of SAECG proposed by Simson [1]).

In this work, *distance based exploratory data analysis* of the ECG waveforms is utilized for the advance of signal averaging. Such analysis includes outlier detection and clustering and its results are expressed via the theory of fuzzy sets. The ECG segments are weighted before averaging according to their membership value.

Two main methods are proposed, that can appear in many variants depending upon the representation of the ECG waveforms. The first is based on the ***Maximum-Likelihood*** (**ML**) theory, while the second on the principles of ***Cluster Analysis*** (**CA**). Both methods involve vectorial consideration of ECG patterns and computation of *fuzzy* controlled weights. The scope of the two Fuzzy Weighted Averaging (**FWA**) techniques is to improve the diagnostability of SAECG by guaranteeing high SNR and enhancing the useful characteristics.

The new averaging techniques are presented in an order according to their complexity and their performance is evaluated using real data.

I. FWA based on ML

After the alignment of the beats [1], a pattern representing each one is selected. This pattern is considered as a multivariate observation (vector):

$$ECG_i(t), t = 1, 2, \ldots \rightarrow \vec{X}_i \quad , i: \#\ of\ heart\ beat \quad (1)$$

This might be a certain segment of the beat around the QRS offset. In that case, components of the vector will be the time instants:

$$\mathbf{X}_i = [\ ECG_i(t_1), \ldots, ECG_i(t_2)\ ].$$

Patterns, based on other representations of the ECG waveforms can also be utilized.

After the *standardization* of the variates [2], distances between vectors can be used as dissimilarity measures. The Euclidean distance of $\mathbf{X}_i$ in respect to the rest of vectors

$$D_i = \sum_{j=1}^{N} \left\| X_i - X_j \right\|_{L_2} \qquad (2)$$

serves as an index for the quality of this vectorial observation. **ML** estimate approach suggests that the smaller the D_i, the more reliable the corresponding $\mathbf{X}_i$ [3]. Outlier vectors correspond to extreme D_i. In that sense, an implicit classification is accomplished.

These concepts are further exploited by assigning a membership value to each vector through a *fuzzy transformation* of the D_i [4]:

$$\mu_i = 1 - \frac{1}{1 + \exp\left(-\frac{D_i - \alpha}{\beta}\right)} \qquad (3)$$

The constant α defines the acceptable range for the D_i in order the corresponding $\mathbf{X}_i$ to be included in the averaging. The β controls the fuzziness of the subset to be averaged and it should be propotional to the noise level as it is measured in the ST segment of the **SAECG** prior the weighting [5].

Detection of the **LP** will be based upon the average of the weighted ECG waveforms

$$Y(t) = \sum_{i=1}^{N} \frac{\mu_i}{\sum \mu_i} ECG_i(t) \qquad (4)$$

The proposed self-adaptive procedure, weights each beat in proportion to its signal content. Movement artifacts or extremely noisy beats are downweighted, while the useful information is marked out.

II. FWA based on CA

Cluster is a set of vectors, where the inter-vector distances are small compared to the distances to vectors outside the cluster. The rationale behind using **CA** is the fact that, inspired beats can deviate a lot from the expired ones in respect of SNR [1]. **CA** is employed to define a tight cluster that includes all the vectors corresponding to high quality ECG waveforms.

The objective is to find those vectors that form the collection with the smallest dispersion:

$$J_{[X_i]} = \frac{1}{I} \sum_{i=1}^{I} \left\| X_i - \overline{X} \right\|^2 , \qquad \text{where } \overline{X} = \frac{1}{I} \sum_{i}^{I} X_i \qquad (5)$$

It can be proved that **J** can be expressed in the form of a summation of pairwise distances:

$$J_{[X_i]} = \frac{1}{I^2} \sum_{i}^{I} \sum_{j>i}^{I} \left\| X_j - X_j \right\|^2 \qquad (6)$$

It is obvious that the minimum of **J** could be found by an exhaustive search in the distance matrix

$$D(i,j) = \left\| X_i - X_j \right\|_{L_2}^2 \qquad (7)$$

Since the computational cost of testing all the possible combinations would be enormous, we adopted a hierarchical scheme summarized in the sequel. The basic idea is that the clustering is improved by including a new vector, only if the dispersion ("variance within the cluster") of the new collection of vectors becomes lower.

THE CLUSTERING ALGORITHM

Find the smallest entry $\mathbf{D(i_{sel},j_{sel})}$ in the distance matrix and start building a cluster, **cluster = [$\mathbf{X_{i\,sel}}$, $\mathbf{X_{j\,sel}}$]** following the steps:

ST(1): Find the vector closer to the already formed cluster, which is the one minimizing

$$f(k) = \sum_{j:\, X_j \in cluster} D(j,k),\; k:\; X_k \notin cluster \quad (8)$$

ST(2): Check if this vector $\mathbf{X_k}$ should be included in the cluster

$$J_{[cluster \cup X_k]} \le J_{[cluster]}, \; where$$

$$J_{[cluster \cup X_k]} = \frac{r^2 \; J_{[cluster]} + f_{\min}}{(r+1)^2} \qquad (9)$$

$$and\; r = \#\; of\; vectors \in cluster$$

ST(3): If dispersion is reduced, include $\mathbf{X_k}$ in the cluster, and repeat from **ST(1)**

A membership value is assigned to each vector through the *fuzzy transformation* of its distance to the cluster centre:

$$\mu_i = 1 - \frac{1}{1 + \exp\left(-\frac{\left\| X_i - O \right\|^2 - \alpha}{\beta}\right)}$$

$$O = \frac{1}{r} \sum_{r} X_j \quad j: X_j \in cluster \;\wedge\; \alpha = J_{cluster}$$

The constant $\boldsymbol{\alpha}$ expresses the acceptable range for $\mathbf{X_i}$ to be included in the cluster, while $\boldsymbol{\beta}$ is set as in equ(3).

LP will be detected in the waveform obtained by equ(4), where the well clustered heart beats are enhanced.

III EXPERIMENTAL RESULTS

For this study 13 normal subjects and 13 postmyocardial infraction patients with sustained Ventricular Tachycardia (**VT**) were included. The signals were recorded from the three bipolar modified Frank leads, using a High Resolution Marquette Case 15 System and digitized at a sampling rate of 2 KHz.

For each class, the signals from 10 subjects were used to construct, via standard signal averaging [1], a set of typical patterns; one for each of the x,y,z leads. The signals from the rest of subjects were used to test the impact of the proposed methods to the classification.

Classification was based on the *Mahalanobis* distance of the waveform under consideration to the two classes centroids [7].

Quantitative results are given in Table I, through the index **P** , where **ì**$_{VT}$ and **ì**$_{NORMAL}$ are the centroids of the two classes and **Y** the waveform for the detection of the **LP**

$$P = \frac{\sum_{j=x,y,z} \left\| Y^j - \mu^j_{VT} \right\|_M}{\sum_{j=x,y,z} \left\| Y^j - \mu^j_{\mathrm{normal}} \right\|_M} \quad (11)$$

Our methods decrease this index for the VT patient and increases it for normal subjects when compared with the conventional Signal Averaging (SA). This is a strong indication that the proposed methods increases the sensitivity and specificity of SAECG.

DISCUSSION

The construction of the pattern (vector) representing each ECG waveform, is an important part of the proposed methods. Features, with diagnostic importance for the examination of SAECG (e.g details D5, D4 and D3 of the Wavelet Transform [6]), can be utilized as components of these vectors. **Our methods provide average-waveforms with these features better portrayed and this could lead to a more reliable diagnosis**. Membership values for each heart beat signal can be derived from distinct approaches and combined through the theory of fuzzy sets [8]. These ideas are under consideration with promising preliminary results.

REFERENCES

[1] Lander P & Berbari E. *Principles and Signal Processing Techniques of the HRECG*.Progress in Cardiovascular Diseases, XXXV(3), 1992:169-188

[2] Bryan F.J. Manly. *"Tests of significance with multivariate data"* in Multivariate Statistical Methods. (ed.) Chapman & Hall

[3] Astola J. et al. *Vector median filters*. Proc. IEEE, 1990, 78: 678-689.

[4] Plataniotis K.N et al. *Colour Image Processing Using Fuzzy Vector Directional filters*. IEEE Workshop on Nonlinear Processing,Halkidiki,Greece 1995,pp535-8

[5] Lander P. et al. *Optimal Filtering and Quality Control of the Signal-Averaged ECG*. Circulation, 1995, 91/5, 1496-1505

[6] Bianchi A. et al. *Detection and Frequency Correction of Ventricular LPs through Wavelet Transform*. Computers in Cardiology, Vienna,Austria 1995, pp75

[7] Cohen A. *"Compression and Automatic Recognition"* in Biomedical Signal Processing vol.II (ed.) CRC Press

[8] Ross T.J. *"Classical Sets and Fuzzy Sets"* in Fuzzy Logic with Engineering Applications. (ed.) McGraw-Hill,Inc

Table I

	SA	FWA_{ML}	FWA_{CA}
VT_1	0.78	0.63	0.52
VT_2	0.85	0.75	0.69
VT_3	0.93	0.88	0.62
$Normal_1$	1.11	1.19	1.22
$Normal_2$	1.06	1.20	1.31
$Normal_3$	1.23	1.26	1.33