SIGNAL DEPENDENT KL TRANSFORM FOR ECG SIGNALS REPRESENTATION

by
Cristin BIGAN

Abstract: The general transform processes are all members of the signal independent, or fixed, class of transforms. All such transforms have a predefined transform vectors which do not change. The advantage of such transforms besides fast algorithms, is the selection of a transform according to its specific properties in relation to a signal of known characteristics. In many instances, however, the characteristics of the signal may be complex, or even unknown. In such cases selection of an unsuitable transform with give poor results. Signals may also contain components suited to different transforms. A signal comprising pure tones plus Dirac functions, for example, will not be efficiently processed by a Fourier transform, which cannot express Dirac functions easily.
The statistically best block transform (in terms of decorrelation of the signal samples, and energy repacking) is, a signal dependent one, the Karhunen Loeve Transform (KLT) used here for large databases of ECG compression and representation.
Key words: *Block transforms, signal dependent, ECG, representation.*

1. INTRODUCTION

Signal dependent transforms are those transforms that generate their transform vectors based on either apriori knowledge of the signal or by analysis of the signal. The advantage of such schemes is that they can adjust themselves to the characteristics of the signal. The main disadvantage is that they tend to be more computationally intensive. This is caused by two factors. Firstly, if the transform vectors are not predefined then a fast algorithm for their use cannot be generated. Secondly, there will be a computational overhead for the generation of the vectors in the first place, which can be very large. There is also the problem of storage of the transform vectors. As these vectors are unique for the data set under analysis they must all be stored. This overhead can become relatively large if the number of waveforms in the set is small, or if the length of each waveform is large.

2. THE KARHUNEN LOEVE TRANSFORM (KLT)

The KLT can be used on any signal that comprises a set of correlated data sequences. For example, in image processing a film sequence comprises a set of correlated images. For ECG processing, the ECG can be expressed as a set of individual ECG waveforms, which is a correlated data set, as shown in Fig.1.

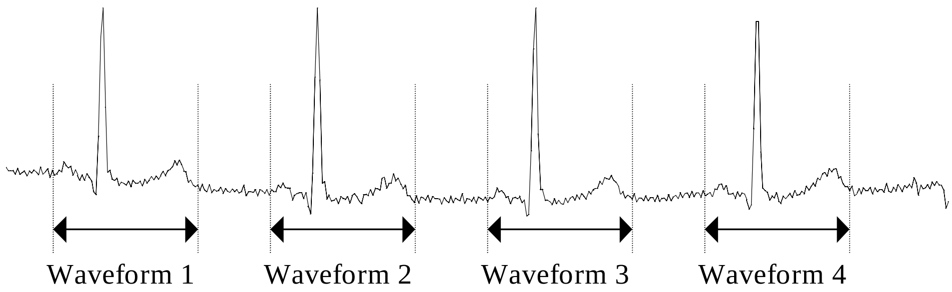


Fig. 1. Segmentation of the ECG Signal.

The KLT is the optimal transform in that:

- It completely decorrelates the original signal. (the transform coefficients are statistically independent for a Gaussian signal).
- It optimises the repacking of the signal energy, such that most of the signal energy is contained in the fewest transform coefficients.
- The total entropy of the signal is minimised.
- For any amount of compression the MSE in the reconstruction is minimised.

Given these abilities, the KLT should be in widespread use. However, there are several disadvantages to using the KLT, the greatest being the computational overhead required to generate the transform vectors. The transform vectors for the KLT are the eigenvectors of the autocovariance matrix formed from the data set. To generate the KLT vectors the procedure is as follows:

1. Take N sequences (ECGs) each of L data points, $X_{[N][L]}$.
2. Construct an average signal M

$$M_j = \frac{1}{N}\sum_{i=0}^{N-1} X_{[i][j]} \qquad j = 0 \ldots L-1 \tag{1}$$

3. Subtract the average signal from the original ensemble

$$Y_{[i]} = X_{[i]} - M \qquad i = 0 \ldots N-1 \tag{2}$$

4. Construct a covariance matrix.

$$\begin{pmatrix} C_{1,1} & C_{1,2} & \cdots & \cdots & C_{1,L} \\ C_{2,1} & C_{2,2} & & & \\ \vdots & \vdots & \ddots & & \\ \vdots & \vdots & & \ddots & \\ C_{L,1} & C_{L,2} & & & C_{L,L} \end{pmatrix}$$

where each value $C_{i,j}$ is given by

$$C_{i,j} = \sum_{p=0}^{N-1} \frac{Y_{[p][i]} \times Y_{[p][j]}}{N-1} \qquad i = 1...L,\ j = 1...L \qquad (3)$$

It is the eigenvectors of this matrix which are used as the basis vectors for the transform.

To extract the eigenvectors from a matrix of rank *L*, requires processing of the order $L^{2.}$ It is this computation that is often quoted (refs) as being too large to be realisable, even for the relatively small values of *L* required for ECGs. Hence, this is why the KLT is not used, except as a benchmark for other faster schemes. However, given the computing power readily available in modern desktop PCs, this computation is no longer such a problem. Fig. 2 gives the processing time taken to extract the eigenvectors from a matrix, with rank *L* from 50 to 330. The algorithm used is running on a standard Pentium 200Mhz PC.

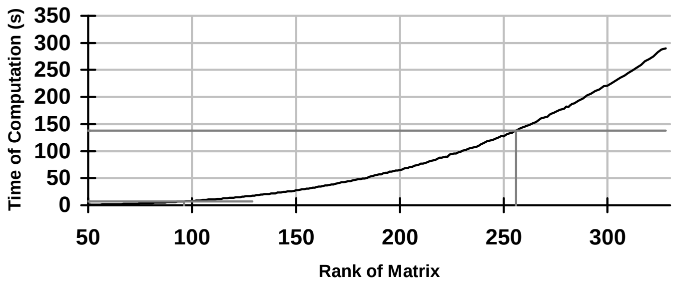


Fig. 2. Computation Time of Eigenvectors.

For the 360Hz ECG recordings from the MIT database the matrix rank is set at 256, and for the 128Hz ECG recordings the rank is set at 96. These are chosen to give data windows of approximately 700ms, and also to maintain compatibility with the wavelet based transform methods. This gives a computation time of 68 seconds and 3.1 seconds respectively to extract the eigenvectors. This is a one off computation for each recording, and so does not appear to be a great overhead. Also this time does not include the additional processing required to fill the covariance matrix, which is of order NL^2.

All the eigenvectors are normalised to have unit energy, to give an orthonormal transform. The corresponding eigenvalues for these vectors show the variance distribution of the ensemble within this transform domain. This indicates how effective the repacking of the signal energy is likely to be with these vectors. Also direct study of the vectors themselves may help in the understanding the mechanisms, and component sources, present in the signal. This can help in either the preprocessing of the signal, such as to eliminate noise sources, or assist in the selection of a faster fixed transform. This is demonstrated in the results below.

3. RESULTS

Consider the signal in Fig.3.

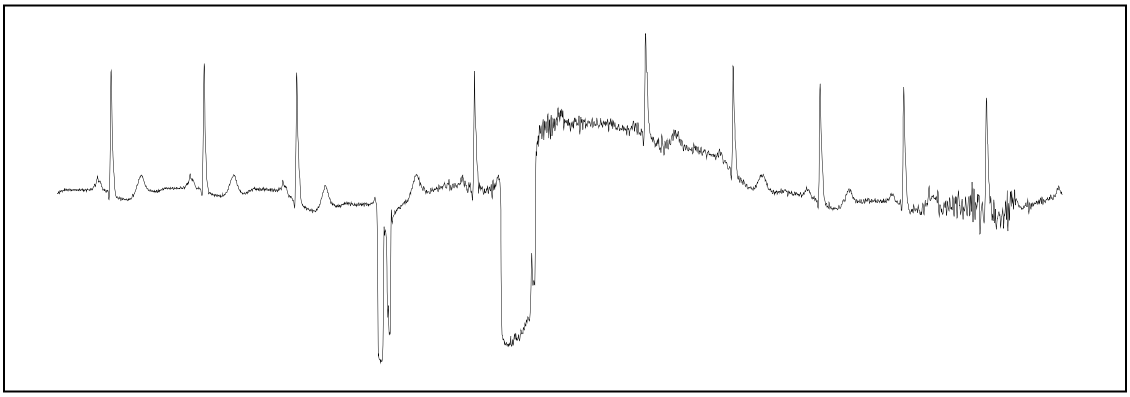


Fig.3. Sample ECG Signal

This is a sample of a recording from the MIT-BIH arrhythmia database (recording number 101, channel 1) well known as a reference data for analyse.

If this tape is used to generate a set of transform vectors, the following results are achieved. Firstly an average waveform is generated, Fig.4, that is then subtracted from all the waveforms.

From the subsequent covariance matrix the eigenvalues and eigenvectors can be extracted. As it is the eigenvalues which express the importance of the eigenvectors, the vectors are ordered in descending order of their corresponding eigenvalues.

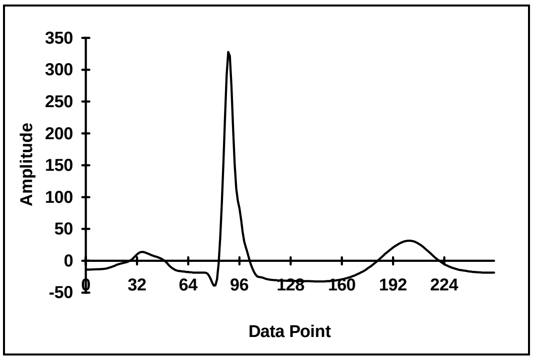


Fig.4. Average Waveform

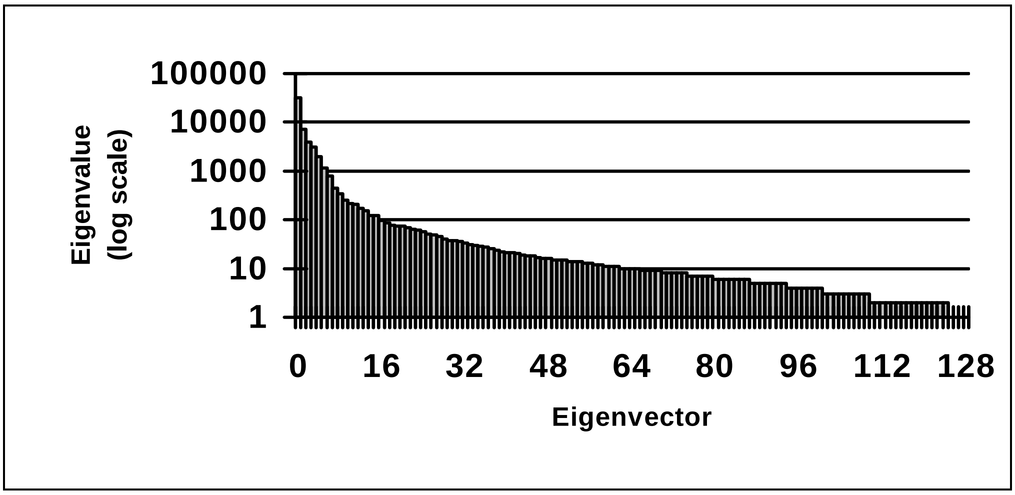
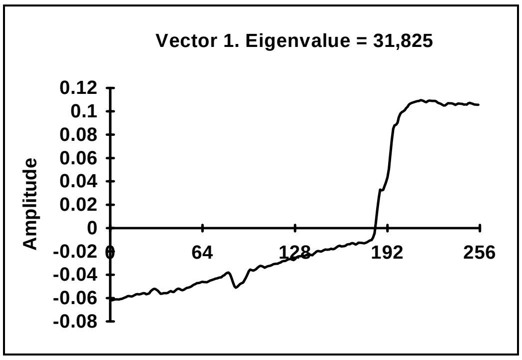
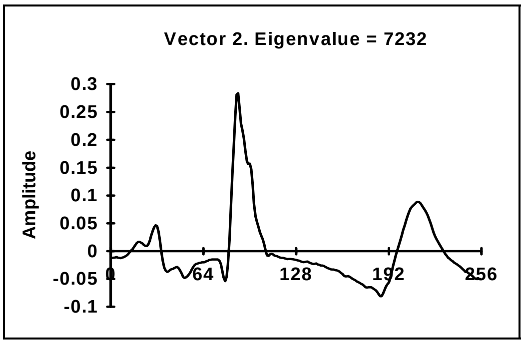
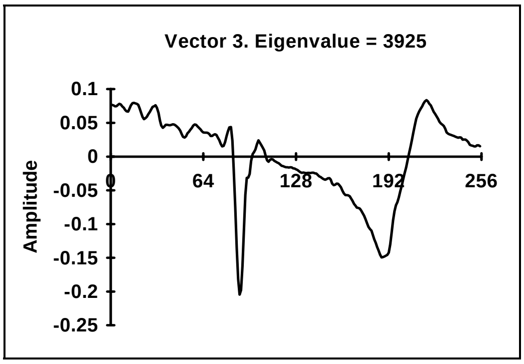


Fig.5. Eigenvalues of the Principal Eigenvectors

The eigenvalues give a measure of how the energy of the transform is distributed, and so indicate how well the energy of the signal will be redistributed in the transform. So above it is clear that this transform should perform well, as the first five eigenvectors of this transform represent 99.8% of the transform energy, which means discarding the rest of the vectors would, on average, lead to an error in the transform of less than 5%, as measured by the PRD. The first five eigenvectors are shown below.


Vector 1. Eigenvalue = 31,825
Amplitude
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
0
64
128
192
256


Vector 2. Eigenvalue = 7232
Amplitude
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
0
64
128
192
256


Vector 3. Eigenvalue = 3925
Amplitude
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
0
64
128
192
256

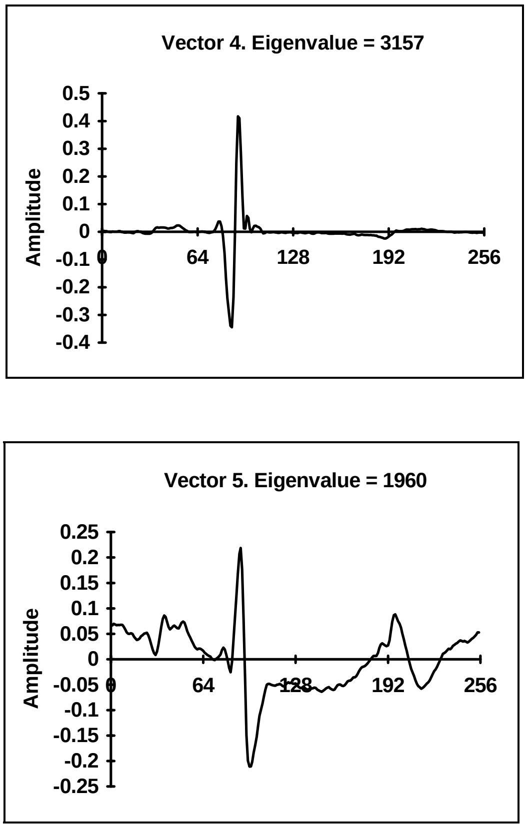


Fig. 6. Principal Eigenvectors

4. CONCLUSIONS

Vector one has the largest eigenvalue and so can be considered to be the most significant. It features an increasing baseline drift with a sudden jump about 190 points into the vector. From the sample of the tape shown above such sudden baseline shifts do occur, although infrequently. However the shifts are of large magnitude, and so have a lot of signal energy associated with them, which is why it has been selected as the most significant feature. All the vectors have some significant feature between points 80 and 100, which is coincident with the QRS complex in the original waveforms. This suggest that the average waveform, which is subtracted from each of the original waveforms, does not remove all the significant features of the QRS complex. This in part could be due to the QRS complex varying over the recording. However it could be caused by the original waveforms being slightly misaligned when

subtracting the average, vector 4 appears to be a differential of the average waveform, which suggests misalignment.
As a main conclusion is the fact that due to the apriori knowledge on the ECG waveform and its content the KL (signal dependent) transform is well suited for such data compression having an average effectiveness upon either block or time-frequency transforms.

REFERENCES

[1]. Blanco, S. Kochen, O.A. Rosso and P. Salgado, *Applying time-frequency analysis to seizure EEG activity*, IEEE Engineering in Medicine and Biology., January-February, 64-71, 1997.
[2]. Olkkonen, .H - *Running discrete Fourier transform for time-frequency analysis of biomedical signals*, Med. Eng.Phys 1995 Vol 17 pp 455-458
[3]. Philips, W - *Adaptive noise removal from biomedical signals using warped polynomials*, IEEE Transactions on Biomedical Engineering - May 1996 Vol- 43 NO- 5 pp- 480-492
[4]. Turbes, C.C.- *Frequency and time domain studies of the micro-EEG from the brain extracellular space* - Biomedical Sciences Instrumentation - 1996 Vol- 32 pp.113-118.
[5]. Fuller W,A, *Introduction to statistical time series*, Wiley Series in Probability and Statistics., 2 Ed., 1996.
[6]. Dripps, J. - *An Introduction to time -frequency methods applied to Biomedical Signals* - in Time-freq. analysis of biomed. signals IEE Colloq. Jan 1997
[7]. Pradhan, N, Dutt, DN, Sadasivan, PK, Ramesh, RG - *Estimation of attractor dimension of EEG using SVD*, Sadhana-Acad. Proceedings in Engineering Sciences -Feb. 1996 Vol- 21 pp- 21-38
[8]. Martens, W. - *The fast time frequency transform : a novel on -line approach to the instantaneous spectrum* in Proceedings of the 14-th International Conference IEEE Eng. in med. and biology 1992 pp. 2594-2596

Author: Cristin BIGAN, Politehnica University of Bucharest, Unirii 15, ap.9 sect. 5 Bucharest, Romania, (00401) 3362939, Email: ibigan@pcnet.pcnet.ro