3

Non-Stationary Biosignal Modelling

Carlos S. Lima, Adriano Tavares, José H. Correia,
Manuel J. Cardoso[1] and Daniel Barbosa
University of Minho
Portugal
[1]University College of London
England

1. Introduction

Signals of biomedical nature are in the most cases characterized by short, impulse-like events that represent transitions between different phases of a biological cycle. As an example hearth sounds are essentially events that represent transitions between the different hemodynamic phases of the cardiac cycle. Classical techniques in general analyze the signal over long periods thus they are not adequate to model impulse-like events. High variability and the very often necessity to combine features temporally well localized with others well localized in frequency remains perhaps the most important challenges not yet completely solved for the most part of biomedical signal modeling. Wavelet Transform (WT) provides the ability to localize the information in the time-frequency plane; in particular, they are capable of trading on type of resolution for the other, which makes them especially suitable for the analysis of non-stationary signals.

State of the art automatic diagnosis algorithms usually rely on pattern recognition based approaches. Hidden Markov Models (HMM's) are statistically based pattern recognition techniques with the ability to break a signal in almost stationary segments in a framework known as quasi-stationary modeling. In this framework each segment can be modeled by classical approaches, since the signal is considered stationary in the segment, and at a whole a quasi-stationary approach is obtained.

Recently Discrete Wavelet Transform (DWT) and HMM's have been combined as an effort to increase the accuracy of pattern recognition based approaches regarding automatic diagnosis purposes. Two main motivations have been appointed to support the approach. Firstly, in each segment the signal can not be exactly stationary and in this situation the DWT is perhaps more appropriate than classical techniques that usually considers stationarity. Secondly, even if the process is exactly stationary over the entire segment the capacity given by the WT of simultaneously observing the signal at various scales (at different levels of focus), each one emphasizing different characteristics can be very beneficial regarding classification purposes.

This chapter presents an overview of the various uses of the WT and HMM's in Computer Assisted Diagnosis (CAD) in medicine. Their most important properties regarding biomedical applications are firstly described. The analogy between the WT and some of the

biological processing that occurs in the early components of the visual and auditory systems, which partially supports the WT applications in medicine is shortly described. The use of the WT in the analyses of 1-D physiological signals especially electrocardiography (ECG) and phonocardiography (PCG) are then reviewed. A survey of recent wavelet developments in medical imaging is then provided. These include biomedical image processing algorithms as noise reduction, image enhancement and detection of micro-calcifications in mammograms, image reconstruction and acquisition schemes as tomography and Magnetic Resonance Imaging (MRI), and multi-resolution methods for the registration and statistical analysis of functional images of the brain as positron emission tomography (PET) and functional MRI.
The chapter provides an almost complete theoretical explanation of HMMs. Then a review of HMMs in electrocardiography and phonocardiography is given. Finally more recent approaches involving both WT and HMMs specifically in electrocardiography and phonocardiography are reviewed.

2. Wavelets and biomedical signals

Biomedical applications usually require most sophisticated signal processing techniques than others fields of engineering. The information of interest is often a combination of features that are well localized in space and time. Some examples are spikes and transients in electroencephalograph signals and microcalcifications in mammograms and others more diffuse as texture, small oscillations and bursts. This universe of events at opposite extremes in the time-frequency localization can not be efficiently handled by classical signal processing techniques mostly based on the Fourier analysis. In the past few years, researchers from mathematics and signal processing have developed the concept of multiscale representation for signal analysis purposes (Vetterli & Kovacevic, 1995). These wavelet based representations have over the traditional Fourier techniques the advantage of localize the information in the time-frequency plane. They are capable of trading one type of resolution for the other, which makes them especially suitable for modelling non-stationary events. Due to these characteristics of the WT and the difficult conditions frequently encountered in biomedical signal analysis, WT based techniques proliferated in medical applications ranging from the more traditional physiological signals such as ECG to the most recent imaging modalities as PET and MRI. Theoretically wavelet analysis is a reasonably complicated mathematical discipline, at least for most biomedical engineers, and consequently a detailed analysis of this technique is out of the scope of this chapter. The interested reader can find detailed references such as (Vetterli & Kovacevic, 1995) and (Mallat, 1998). The purpose of this chapter is only to emphasize the wavelet properties more related to current biomedical applications.

2.1 The wavelet transform - An overview

The wavelet transform (WT) is a signal representation in a scale-time space, where each scale represents a focus level of the signal and therefore can be seen as a result of a band-pass filtering.
Given a time-varying signal x(t), WTs are a set of coefficients that are inner products of the signal with a family of *wavelets* basis functions obtained from a standard function known as

mother wavelet. In Continuous Wavelet Transform (CWT) the wavelet corresponding to scale s and time location τ is given by

$$\psi_{\tau,s} = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \tag{1}$$

where ψ(t) is the mother wavelet, which can be viewed as a band-pass function. The term $\sqrt{|s|}$ ensures energy preservation. In the CWT the time-scale parameters vary continuously. The wavelet transform of a continuous time varying signal x(t) is given by

$$\Psi_x^{\psi}(\tau, s) = \frac{1}{\sqrt{|s|}}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-\tau}{s}\right)dt \tag{2}$$

where the asterisk stands for complex conjugate. Equation (2) shows that the WT is the convolution between the signal and the wavelet function at scale s. For a fixed value of the scale parameter s, the WT which is now a function of the continuous shift parameter τ, can be written as a convolution equation where the filter corresponds to a rescaled and time-reversed version of the wavelet as shown by equation (1) setting t=0. From the time scaling property of the Fourier Transform the frequency response of the wavelet filter is given by

$$\frac{1}{\sqrt{|s|}}\psi\left(-\frac{\tau}{s}\right) \qquad \leftrightarrow \qquad \sqrt{|s|}\Psi^*(s\omega) \tag{3}$$

One important property of the wavelet filter is that for a discrete set of scales, namely the dyadic scale $s = 2^i$ a constant-Q filterbank is obtained, where the quality factor of the filter is defined as the central frequency to bandwidth ratio. Therefore WT provides a decomposition of a signal into subbands with a bandwidth that increases linearly with the frequency. Under this framework the WT can be viewed as a special kind of spectral analyser. Energy estimates in different bands or related measures can discriminate between various physiological states (Akay & al. 1994). Under this approach, the purpose is to analyse turbulent hearth sounds to detect coronary artery disease. The purpose of the approach followed by (Akay & Szeto 1994) is to characterize the states of fetal electrocortical activity. However, this type of global feature extraction assumes stationarity, therefore similar results can also be obtained using more conventional Fourier techniques. Wavelets viewed as a filterbank have motivated several approaches based on reversible wavelet decomposition such as noise reduction and image enhancement algorithms. The principle is to handle selectively the wavelet components prior to reconstruction. (Mallat & Zhong, 1992) used such a filterbank system to obtain a multiscale edge representation of a signal from its wavelets maxima. They proposed an iterative algorithm that reconstructs a very close approximation of the original from this subset of features. This approach has been adapted for noise reduction in evoked response potentials and in MR images and also in image enhancement regarding the detection of microcalcifications in mammograms.

From the filterbank point of view the shape of the mother wavelet seems to be important in order to emphasize some signal characteristics, however this topic is not explored in the ambit of the present chapter.
Regarding implementation issues both s and τ must be discretized. The most usual way to sample the time-scale plane is on a so-called *dyadic* grid, meaning that sampled points in the time-scale plane are separated by a power of two. This procedure leads to an increase in computational efficiency for both WT and Inverse Wavelet Transform (IWT). Under this constraint the Discrete Wavelet Transform (DWT) is defined as

$$\psi_{j,k}(t) = s_0^{-j/2}\psi\left(s_0^{-j}t - k\tau_0\right) \tag{4}$$

which means that DWT coefficients are sampled from CWT coefficients. As a dyadic scale is used and therefore $s_0=2$ and $\tau_0=1$, yielding $s=2^j$ and $\tau=k2^j$ where j and k are integers.
As the scale represents the level of focus from the which the signal is viewed, which is related to the frequency range involved, the digital filter banks are appropriated to break the signal in different scales (bands). If the progression in the scale is *dyadic* the signal can be sequentially half-band high-pass and low-pass filtered.

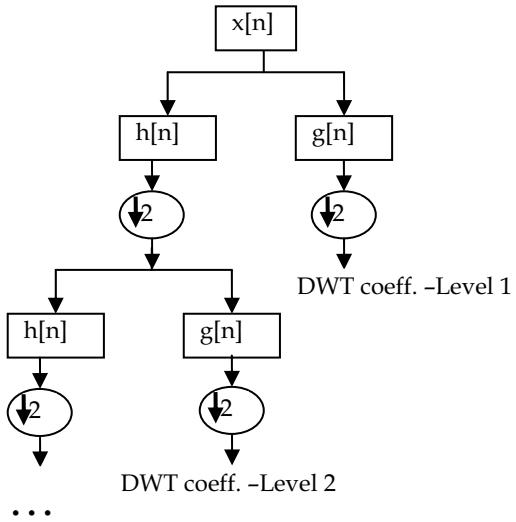

Fig. 1. Wavelet decomposition tree

The output of the high-pass filter represents the detail of the signal. The output of the low-pass filter represents the approximation of the signal for each decomposition level, and will be decomposed in its detail and approximation components at the next decomposition level. The process proceeds iteratively in a scheme known as wavelet decomposition tree, which is

shown in figure 1. After filtering, half of the samples can be eliminated according to the Nyquist's rule, since the signal now has only half of the frequency.
This very practical filtering algorithm yields as Fast Wavelet Transform (FWT) and is known in the signal processing community as two-channel subband coder.
One important property of the DWT is the relationship between the impulse responses of the high-pass (g[n]) and low-pass (h[n]) filters, which are not independent of each other and are related by

$$g[L-1-n] = (-1)^n h[n] \tag{5}$$

where L is the filter length in number of points. Since the two filters are odd index alternated reversed versions of each other they are known as Quadrature Mirror Filters (QMF). Perfect reconstruction requires, in principle, ideal half-band filtering. Although it is not possible to realize ideal filters, under certain conditions it is possible to find filters that provide perfect reconstruction. Perhaps the most famous were developed by Ingrid Daubechies and are known as Daubechies' wavelets. This processing scheme is extended to image processing where temporal filters are changed by spatial filters and filtering is usually performed in three directions; horizontal, vertical and diagonal being the filtering in the diagonal direction obtained from high pass filters in both directions.
Wavelet properties can also be viewed as other approaches than filterbanks. As a multiscale matched filter WT have been successful applied for events detection in biomedical signal processing. The matched filter is the optimum detector of a deterministic signal in the presence of additive noise. Considering a measure model $f(t) = \psi_s(t - \Delta t) + n(t)$ where $\psi_s(t) = \psi(t/s)$ is a known deterministic signal at scale *s*, Δt is an unknown location parameter and *n(t)* an additive white Gaussian noise component. The maximum likelihood solution based on classical detection theory states that the optimum procedure for estimating Δt is to perform the correlations with all possible shifts of the reference template (convolution) and to select the position that corresponds to the maximum output. Therefore, using a WT-like detector whenever the pattern that we are looking for appears at various scales makes some sense.
Under correlated situations a pre-whitening filter can be applied and the problem can be solved as in the white noise case. In some noise conditions, specifically if the noise has a fractional Brownian motion structure then the wavelet-like structure of the detector is preserved. In this condition the noise average spectrum has the form $N(w) = \sigma^2 / |w|^{\alpha}$ with $a=2H+1$ with *H* as the Hurst exponent and the optimum pre-whitening matched filter at scale *s* as

$$(-j)^{\alpha} D^{\alpha} \psi_s(t) = C_s \psi\left(t/s\right) \tag{6}$$

where D^{α} is the αth derivative operator which corresponds to $(jw)^{\alpha}$ in the Fourier domain. In other words, the real valued wavelet $\psi(t)$ is proportional to the fractional derivative of the pattern ψ that must be detected. For example the optimal detector for finding a Gaussian in $O(w^{-2})$ noise is the second derivative of a Gaussian known as Mexican hat

wavelet. Several biomedical signal processing tasks have been based on the detection properties of the WT such as the detection of interictal spikes in EEG recordings of epileptic patients or cardiology based applications as the detection of the QRS complex in ECG (Li & Zheng, 1993). This last application also exploits the ability of the WT to characterize singularities through the decay of the wavelet coefficients across scale. Detection of microcalcifications in mammograms is another application that successfully uses the detection properties of the WT (Strickland & Hahn, 1994).

2.2 2D Wavelet Transform

The reasoning explained in section 2.1 can be extended to the bi-dimensional space and applied to image processing. Mallat (Mallat 1989) introduced a very elegant extension of the concepts of multi-resolution decomposition to image processing. The proposed key idea is to expand the application of 1D filterbanks to the 2D in straightforward manner, applying the designed filters to the columns and to the rows separately. The orthogonal wavelet representation of an image can be described as the following recursive convolution and decimation

$$A_n(i,j) = [H_c * [H_r * A_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \qquad D_{n1}(i,j) = [H_c * [G_r * A_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2}$$

$$D_{n2}(i,j) = [G_c * [H_r * A_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \qquad D_{n3}(i,j) = [G_c * [G_r * A_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \tag{7}$$

where $(i,j) \in R^2$, $*$ denotes the convolution operator, $\downarrow 2,1$ ($\downarrow 1,2$) sub-sampling along the rows (columns) and $A_0 = I(x,y)$ is the original image. H and G are low and band pass quadrature mirror filters, respectively. A_n is obtained by low pass filtering leading to a less detailed/approximation image, at scale n. The D_{ni} are obtained by band pass filtering in a specific direction, therefore encoding details in different directions. Thus these parameters contain directional detail information at scale n. This recursive filtering is no more than the extension of the scheme represented in figure 1 to a bi-dimensional space as shown in figure 2.

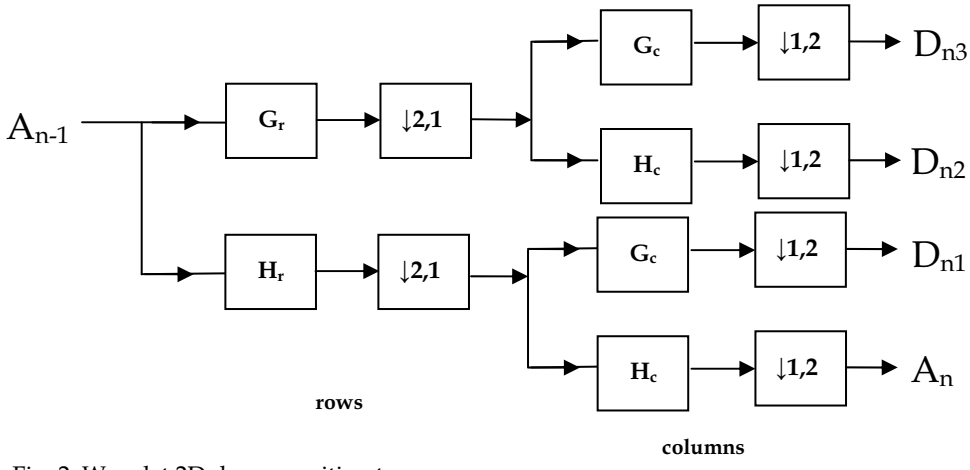

Fig. 2. Wavelet 2D decomposition tree

This 2D implementation is therefore a recursive one-dimensional convolution of the low and band pass filters with the rows and columns of the image, followed by the respective subsampling. One can note that the 2D DWT decomposition is the result at each considered scale, in subbands of different frequency content or detail, in the different orientations. A good example is illustrated in figure 3.

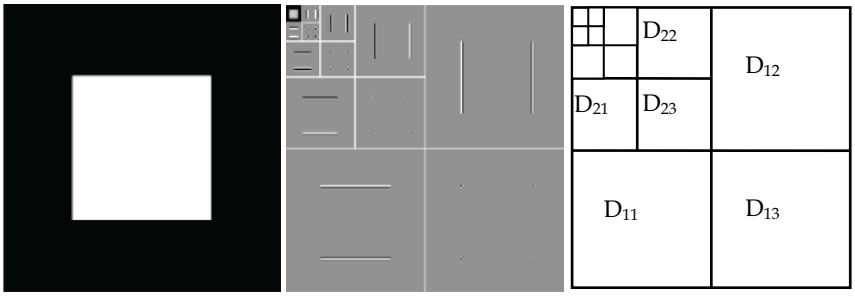

Fig. 3. Decomposition of 2D DWT in sub-bands

The application of a 2D DWT decomposition to an image of N by N pixels returns N by N wavelet coefficients, being therefore a compact representation of the original image. Furthermore, the key information will be sparsely represented, which will be the driving force for compression schemes based on DWT. The reconstruction of the image is possible through the application of the previous filterbank in the opposite direction.

2.3 Time-Frequency Localization and Wavelets

Most biomedical signals of interest include a combination of impulse-like events such as spikes and transients and also more diffuse oscillations such as murmurs and EEG waveforms which may all convey important information for the clinician and consequently regarding automatic diagnosis purposes. Classical methods based on Short Time Fourier Transform (STFT) are well adapted for the later type of events but are much less suited for the analysis of short duration pulses. Hence when both types of events are present in the data the STFT is not completely adequate to offer a reasonable compromise in terms of localization in time and frequency. The main difference of STFT and WT is that in the latter the size of the analysis window is not constant. It varies in inverse proportion of the frequency so that $s = w_0 / w$ where w_0 is the central wavelet frequency. This property enables the WT to zoom in on details, but at the expense of a corresponding loss in spectral resolution. This trade off between localization in time and localization in frequency represents the well known uncertainty principle. In this the name time-frequency analysis corresponds to the trade off between time and space to achieve a better adaptation to the characteristics of the signal.
The Morlet or Gabor wavelet given by

$$\psi(t) = e^{jw_0 t} e^{-t^2/2} \tag{8}$$

has the best time-frequency localization in the sense of the uncertainty principle since the standard deviation of its Gaussian envelope is σ=s. Its Fourier transform is also a Gaussian function with a central frequency $w = w_0 / s$ and a standard deviation $\sigma_w = 1/s$. Thus each analysis template tends to be predominantly located in a certain elliptical region of the time frequency plane. The same qualitative behaviour also applies for other nongaussian wavelet functions. The area of these localization regions is the same for all templates and is constrained by the uncertainty principle as shown in figure 4.

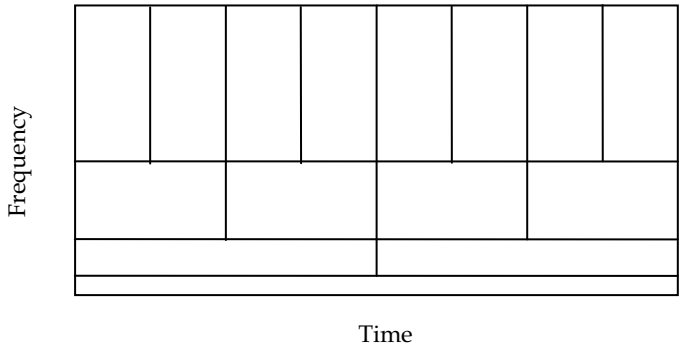

Fig. 4. Time-frequency resolution of the WT

Thus a characterization of the time frequency content of a signal can be obtained by measuring the correlation between the signal and each wavelet template. This reasoning can be extended to image processing where time is replaced by space.
Time frequency wavelet analysis have been used in the characterization of heart beat sounds (Khadra et al.1991, Obaidat 1993, Debbal & Bereksi-Reguig 2004, Debbal & Bereksi-Reguig 2007), the analysis of ECG signals including the detection of late ventricular potentials (Khadra et al. 1993, Dickhaus et al. 1994, Senhadji et al. 1995), the analysis of EEG's (Schiff et al. 1994, Kalayci & Ozdamar 1995) as well as a variety of other physiological signals (Sartene et al. 1994).

2.4 Perception and Wavelets

It is interesting to note that the WT and some of the biological information processing occurring in the first stages of the auditory and visual perception systems are quite similar. This similarity supports the use of wavelet derived methods for low-level auditory and visual sensory processing (Wang & Shamma 1995, Mallat 1989).
Regarding auditory systems, the analysis of acoustic signals in the brain involves two main functional components: 1) the early auditory system which includes the outer ear, middle ear, inner ear or the cochlea and the cochlear nucleus and 2) the central auditory system, which consists of a highly organized neural network in the cortex. Acoustic pressures impinging the outer ear are transmitted to the inner ear, transduced into neural electrical impulses, which are further transformed and processed in the central auditory system. The analysis of sounds in the early and central systems involves a series of processing stages that behave like WT's. In particular it is well known that the cochlea transforms the acoustic pressure *p(t)* received from the middle ear into displacements *y(t,x)* of its basilar membrane

given by $y(t,x)=p(t) * h(t,x)$ where x is the curvilinear coordinate along the cochlea, $h(t,x)=h(ct/x)$ is the cochlear band-pass filter located at x and c the propagation velocity (Yang et al. 1992, Wang & Shamma 1995). Hence $y(t,x)$ is simply the CWT of $p(t)$ with the wavelet $h(t)$ at a time scale proportional to the position x/c. New Engineering applications for the detection, transmission and coding of auditory signals has been inspired in this WT property (Benedetto & Teolis 1993).

Also the visual system includes, among other complex functional units, an important population of neurons that have wavelet-like properties. These are the so-called simple cells of the occipital cortex, which receive information from the retina through the lateral geniculate nucleus and send projections to the complex and hypercomplex cells of the primary and associative visual cortices. Simple cortical cells have been characterized by their frequency response which is a directional bandpass, with a radial bandwidth almost proportional to the central frequency (constant-Q analysis) (Valois & Valois 1988). Topographically, these neurons are organized in such a way that a common preferential orientation is shared, which is not unlike wavelet channels. The receptive fields of these cells, which is the corresponding area on the retina that produces a response, consist of distinct elongated excitatory and inhibitory zones of a given size and orientation being their response approximately linear (Hubel 1982). The spatial responses of individual cells are well represented by modulated Gaussians (Marcelja 1980). Based on these properties, a variety of multichannel neural models consisting of a set of directional Gabor filters with a hierarchical wavelet based organization have been formulated (Daugman 1988, Daugman 1989, Porat & Zeevi 1989, Watson 1987). Simpler decompositions wavelet based analyses have also been considered (Gaudart et al. 1993).

2.5 Wavelets and Bioacoustics

Vibrations caused by the contractile activity of the cardiohemic system generate a sound signal if appropriate transducers are used. The phonocardiogram (PCG) represents the recording of the heart sound signal and provides an indication of the general state of the heart in terms of rhythm and contractility. Cardiovascular diseases and defects can be diagnosed from changes or additional sounds and murmurs present in the PCG. Sounds are short, impulse-like events that represent transitions between the different hemodynamic phases of the cardiac cycle. Murmurs, which are primarily caused by blood flow turbulence, are characteristic of cardiac disease such as valve defects. Given its properties the WT appears to be an appropriate tool for representing and modeling the PCG. A comparative study with other time-frequency methods (Wigner distribution and spectrogram) confirmed its adequacy for this particular application (Obaidat 1993). In particular, certain sound components such as the aortic (A2) and pulmonary (P2) valve components of the second heart sound are hardly resolved by the other methods rather than WT. More recent wavelet based approaches have considered the identification of the two major sounds and murmurs (Chebil & Al-Nabulsi 2007) and also the identification of the components of the second cardiac sound S2 (Debbal & Bereksi-Reguig 2007). Both are of utmost importance regarding diagnosis purposes. In the first case a performance of about 90% is reported which can constitute a very promising result given the difficult conditions existing in situations of severe murmurs. Particularly important in the scope of this chapter is the second situation where the objectives are to determine the order of the closure of the aortic (A2) and pulmonary (P2) valves as well as the time between these two events known as *split*. The

second heart sound S2 can be used in the diagnosis of several heart diseases such as pulmonary valve stenosis and right Bundle branch block (wide split), atrial septal defect and right ventricular failure (fixed split), left bundle branch block (paradoxical or reverse split), therefore it has long been recognized, and its significance is considered by cardiologists as the "key to auscultation of the heart". However the *split* has durations from around 10 ms to 60 ms, making the classification by the human ear a very hard task (Leung et al. 1998). So, an automated method capable of measuring S2 *split* is desirable. However S2 is very hard to deal with since two very similar components (A2 and P2) must be recognized. A2 has often higher amplitude (louder) and frequency content than P2 and generally A2 precedes P2. Several approaches have been proposed to face this problem. In the ambit of this chapter we will focus on the WT since other methods can not resolve the aortic and pulmonary components as stated by (Obaidat 1993). (Debbal & Bereksi-Reguig 2007) proposed an interesting approach entirely based on WT to segment the heart sound S2. Very promising results were obtained by decomposing S2 into a number of components using the WT and chose two of the major components as A2 and P2 in order to define the *split* as the time between these components. However the method suffers from an important drawback; since the amplitudes of A2 and P2 are significantly affected by the recording locations on the chest, the two highest components obtained from WT might not always represent A2 and P2. These are strong requirements regarding diagnosis purposes that claim for high accurate measures.

Alternative methods based also on time-frequency representation by using the Wigner Ville distribution of S2 have been suggested (Xu et al. 2000, Xu et al. 2001). However the masking operation which is central to the procedure is done manually making the algorithm very sensitive to errors while performing the masking operation. This happens because A2 and P2 are reconstructed from masked time-frequency representation of the signal. Recent advances in the scope of this approach focus on the Instantaneous Frequency (IF) trajectory of S2 (Yildirim & Ansari 2007). The IF trace was analyzed by processing the data with a frequency-selective differentiator which preserves the derivative information for the spectral components of the IF data of interest. The zero crossings are identified to locate the onset of P2. While this approach appears to be robust against changes in sensor placement, since it relies only in the spectral content of the signal and not also in its magnitude, the performance of the algorithm remains to be validated. As a matter of fact murmurs change the spectral content of the signal and can compromise the algorithm performance.

Although approaches that rely on the separation of A2 and P2 are in general more susceptible to noise and sensor placement conditions robust methods based on Blind Source Separation (BSS) have also been proposed to estimate the split by separating A2 and P2 (Nigam & Priemer 2006). The main criticism of this approach is related with the independency supposition. Since A2 is generated by the closure of the valve between left ventricular and aorta and P2 by the closure of the valve between right ventricular and pulmonic artery, it is very unlikely that an abnormality in the left ventricle does not affect right ventricle too. Hence the assumption of independence between A2 and P2 needs to be validated.

High accuracy methods such as Hidden Markov Models with features extracted from WT can be more adequate than WT alone to model the phonocardiogram, especially if the wave separation is not required for training purposes. Each event (M1, T1, A2, P2 and background) is modeled by its own HMM and training can be done by HMM concatenation

according to the labeling file prepared by the physician (Lima & Barbosa 2008). The order of occurrence of A2 and P2 can be obtained by the likelihood of both hypothesis (A2 preceding P2 and vice versa) and the *split* can be estimated by the backtracking procedure in the Viterbi algorithm which gives the most likely state sequence.

2.6 Wavelets and the ECG

A number of wavelet based techniques have recently been proposed to the analysis of ECG signals. Subjects as timing, morphology, distortions, noise, detection of localized abnormalities, heart rate variability, arrhythmias and data compression has been the main topics where wavelet based techniques have been experimented.

2.6.1 Wavelets for ECG delineation

The time varying morphology of the ECG is subject to physiological conditions and the presence of noise seriously compromise the delineation of the electrical activity of the heart. The potential of wavelet based feature extraction for discriminating between normal and abnormal cardiac patterns has been demonstrated (Senhadji et al., 1995). An algorithm for the detection and measurement of the onset and the offset of the QRS complex and P and T waves based on modulus maxima-based wavelet analysis employing the dyadic WT was proposed (Sahambi et al., 1997a and 1997b). This algorithm performs well in the presence of modeled baseline drift and high frequency additive noise. Improvements to the technique are described in (Sahambi et al., 1998). Launch points and wavelet extreme were both proposed to obtain reliable amplitude and duration parameters from the ECG (Sivannarayana & Reddy 1999).

QRS detection is extremely useful for both finding the fiducial points employed in ensemble averaging analysis methods and for computing the R-R time series from which a variety of heart rate variability (HRV) measures can be extracted. (Li et al., 1995) proposed a wavelet based QRS detection method based on finding the modulus maxima larger than an updated threshold obtained from the preprocessing of pre-selected initial beats. Performances of 99.90% sensitivity and 99.94% positive predictivity were reported in the MIT-BIH database. Several Algorithms based on (Li et al., 1995) have been extended to the detection of ventricular premature contractions (Shyu et al., 2004) and to the ECG robust delineation (Martinez et al., 2004) especially the detection of peaks, onsets and offsets of the QRS complexes and P and T waves.

Kadambe et al., 1999) have described an algorithm which finds the local maxima of two consecutive dyadic wavelet scales, and compared them in order to classify local maxima produced by R waves and noise. A sensitivity of 96.84% and a positive predictivity of 95.20% were reported. More recently the work of (Li et al. 1995) and (Kadambe et al. 1999) have been extended (Romero Lagarreta et al., 2005) by using the CWT, which affords high time-frequency resolution which provides a better definition of the QRS modulus maxima lines to filter out the QRS from other signal morphologies including baseline wandering and noise. A sensitivity of 99.53% and a positive predictivity of 99.73% were reported with signals acquired at the Coronary Care Unit at the Royal Infirmary of Edinburgh and a sensitivity of 99.70% and a positive predictivity of 99.68% were reported in the MIT-BIH database.

Wavelet based filters have been proposed to minimize the wandering distortions (Park et al., 1998) and to remove motion artifacts in ECG's (Park et al., 2001). Wavelet based noise reduction methods for ECG signals have also been proposed (Inoue & Miyazaki 1998, Tikkanen 1999). Other wavelet based denoising algorithms have been proposed to remove the ECG signal from the electrohysterogram (Leman & Marque 2000) or to suppress electromyogram noise from the ECG (Nikoliaev et al., 2001).

2.6.2 Wavelets and arrhythmias

In some applications the wavelet analysis has shown to be superior to other analysis methods (Yi et al. 2000). High performances have been reported (Govindan et al. 1997, Al-Fahoum & Howitt 1999) and new methods have been developed and implemented in implantable devices (Zhang et al. (1999). One approach that combines WT and radial basis functions was proposed (Al-Fahoum & Howitt 1999) for the automatic detection and classification of arrhythmias where the Daubechies D4 WT is used. High scores of 97.5% correct classification of arrhythmia with 100% correct classification for both ventricular fibrillation and ventricular tachycardia were reported. (Duverney et al. 2002) proposed a combined wavelet transform-fractal analysis method for the automatic detection of atrial fibrillation (AF) from heart rate intervals. AF is associated with the asynchronous contraction of the atrial muscle fibers is the most prevalent cardiac arrhythmia in the west world and is associated with significant morbidity. Performances of 96,1% of sensitivity and 92.6% specificity were reported.

Human Ventricular Fibrillation (VF) wavelet based studies have demonstrated that a rich underlying structure is contained in the signal, however hidden to classical Fourier techniques, contrarily to the previous thought that this pathology is characterized by a disorganized and unstructured electrical activity of the heart (Addison et al., 2000, Watson et al., 2000). Based on these results a wavelet based method for the prediction of the outcome from defibrillation shock in human VF was proposed (Watson et al., 2004). An enhanced version of this method employing entropy measures of selected modulus maxima achieves performances of over 60% specificity at 95% sensitivity for predicting a return of spontaneous circulation. The best of alternative techniques based on a variety of measures including Fourier, fractal, angular velocity, etc typically achieves 50% specificity at 95% sensitivity. This enhancement is due to the ability of the wavelet transform to isolate and extract specific spectral-temporal information. The incorporation of such outcome prediction technologies within defibrillation devices will significantly alter their function as current standard protocols, involving sequences of shocks and CPR, which can be altered according on the likelihood of success of a shock. If the likelihood of success is low an alternative therapy prior to shock will be used.

2.7 Wavelets and Medical Imaging

The impact of the Wavelet Transform in the research community is well perceived through the amount of papers and books published since the milestone works of Daubechies (Daubechies 1988) and Mallat (Mallat 1989). Accordingly with Unser (Unser 2003), more than 9000 papers and 200 books were published between the late eighties and 2003, with a significant part being focused in biomedical applications. The first paper describing a medical application of wavelet processing appeared in 1991, where was proposed a

denoising algorithm based in soft-thresholding in the wavelet domain by Weaver et al. (Weaver 1991). Without the claim of being exhaustive, the main applications of wavelets in medical imaging have been:

Image denoising – The multi-scale decomposition of the DWT offers a very effective separation of the spectral components of the original image. The most tipycal denoising strategy takes advantage of this property to select the most relevant wavelet coefficients applying thresholding techniques. Some classic examples of this approach are given in (Jin 2004).

Compression of medical images – The evolution in medical imaging technology implies a fast pace increase in the amount of data generated in each exam, which generate a huge pressure in the storage and networking information systems, being therefore imperative to apply compression strategies. However the compression of medical image is a very delicate subject, since discarding small details may lead to misevaluation of exams, causing severe human and legal consequences (Schelkens 2003). Nevertheless, it should be noted that the sparse representation of the image content given by the DWT coefficients allows the implementation of different compression algorithms, that can go from a lossy compression, with very high compression ratios, to more refined, lossless compression schemes, with minimal loss of information.

Wavelet-based feature extraction and classification – The wavelet decomposition of an image allows the application of different pattern analysis techniques, since the image content is subdivided into different bands of different frequency and orientation detail. Some of the more notable applications have been the texture features extraction from the DWT coefficients, which has been successfully applied in the medical field for abnormal tissue classification (Karkanis 2003, Barbosa et al. 2008, Lima et al. 2008), given that texture can be roughly described as a spatial pattern of medium to high frequency, where the relationship of the pixels within an neighborhood presents different frequencies at different orientations, which can be modeled by the 2D DWT of the image. The use of wavelet features has also been vastly explored in the classification of mammograms, given that different wavelet approaches may be customized in order to better detect suspicious area. These are normally microcalcifications, which are believed to be cancer early indicators, and correspond to bright spots in the image, being usually detected as high frequency objects with small dimensions within the image. Some examples of this application are the works of Lemaur (Lemaur 2003) and Sung-Nien (Sung-Nien 2006).

Tomographic reconstruction – Tomography medical modalities like CT, SPECT or PET gather multiple projections of the human body that have to be reconstructed from the acquired signal, the sinogram. Therefore rely on an instable inverse problem of spatial signal reconstruction from sampled line projections, which is usually done through back projection of the sinogram signal via Radon transform and regularization for removal of noisy artifacts.

This regularization can be improved through the use of wavelet thresholding estimators (Kalifa 2003). Jin et al. (Jin 2003) proposed the noise reduction in the reconstructed through cross-regularization of wavelet coefficients.

Wavelet-encoded MRI – Wavelet basis can be used in MRI encoding schemes, taking advantage from the better spatial localization when compared with the conventional phase-encoded MRI, which uses Fourier basis. This fact allows faster acquisitions than the conventional phase encoding techniques but it is still slower than echo planar MRI (Unser 1996).

Image enhancement – Medical imaging modalities with reduced contrast may require the application of image enhancement techniques in order to improve the diagnostic potential. A typical example is the mammography, where the contrast between the target objects and the soft tissues of the breast is inherently. The easiest approach uses a philosophy similar to the image denoising techniques, where in this case instead of suppressing the unwanted wavelet coefficients one should amplify the interesting image features. Given the original data quality, redundant wavelet transforms are usually used in enhancement algorithms. Examples of enhancement algorithms using wavelets are presented in (Heinlein et al. 2003, Papadopoulos et al. 2008, Przelaskowski et al. 2007).

2.8 Breaking the limits of the DWT

The multi-resolution capability of the DWT has been vastly explored in several fields of signal and image processing, as seen in the last section. The ability of dealing with singularities is another important advantage of the DWT, since wavelets provide and optimal representation for one-dimensional piecewise smooth signal (Do 2005). However natural images are not simply stacks of 1-D piecewise smooth scan-lines, and therefore singularities points are usually located along smooth curves. The DWT inability while dealing with intermediate dimensional structures like discontinuities along curves (Candès 2000) is easily comprehensible, since its directional sensitivity is limited to three directions. Given that such discontinuity elements are vital in the analysis of any image, including the medical ones, a vigorous research effort has been exerted in order to provide better adapted alternatives by combining ideas from geometry with ideas from traditional multi-scale analysis (Candès 2005). Therefore, and as it was realized that Fourier methods were not good for all purposes, the limitations of the DWT triggered the quest for new concepts capable of overcome these limits.

Given that the focus of the present chapter is not the limits of the DWT itself, only a brief overview regarding multi-directional and multi-scale transforms will be given. The steerable pyramids, proposed in the early nineties (Simoncelli 1992, Simoncelli 1995), was one of the first approaches to this problem, being a practical, data-friendly strategy to extract information at different scales and angles. More recently, the curvelet transform (Candès 2000) and the contourlet transform (Do 2005) have been introduced, being exciting and promising new image analysis techniques whose application to medical image is starting to prove its usefulness.

Originally introduced in 2000, by Candès and Donoho, the continuous curvelet transform (CCT) is based in an anisotropic notion of scale and high directional sensitivity in multiple directions. Contrarily to the DWT bases, which are oriented only in the horizontal, vertical and diagonal directions in consequence to the previously explained filterbank applied in the 2D DWT, the elements in the curvelet transform present a high directional sensitivity, which results from the anisotropic notion of scale of this tool. The CCT is based in the tilling of the 2D Fourier space in different concentric coronae, one of each divided in a given number of angles, accordingly with a fixed relation, as can be seen in figure 5.

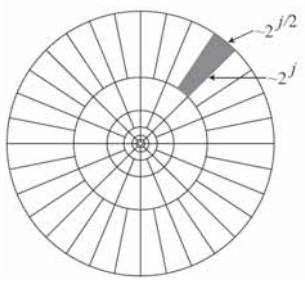

Fig. 5. Tiling of the frequency domain in the continuous curvelet transform

These polar wedges can be defined by the superposition of a radial window $W(r)$ and an angular window $V(t)$. Each of the separated polar wedges will be associated a frequency window U_j, which will correspond to the Fourier transform of a curvelet function $\varphi_j(x)$ function, which can be thought of as a "mother" curvelet, since all the curvelets at scale 2^j may be obtained by rotations and translations of $\varphi_j(x)$. The curvelets coefficients, at a given scale j and angle θ, will be then simply defined as the inner product between the image and the rotation of the mother curvelet φj(x).

Although a discretization scheme has been proposed with its introduction, its complexity was not very user friendly, which led to a redesign of the discretization strategy introduced in (Candès 2006). Nevertheless, the curvelet transform is a concept focused in the continuous domain and has to be discretized to be useful in image processing, given the discrete nature of the pixel grids. This fact has been the seed in (Do & Vetterli 2005), where is proposed a framework for the development of a discrete tool having the desired multi-resolution and directional sensitivity characteristics.

The contourlet tranforms is formulated as a double filter bank, where a Laplacian pyramid is first used to separate the different detail levels and to capture point discontinuities then followed by a directional filter bank to link point discontinuities into linear structures. Therefore the contourlet transform provides a multiscale and directional decomposition in the frequency domain, as can be seen in figure 6, where is clear the division of the Fourier plane by scale and angle.

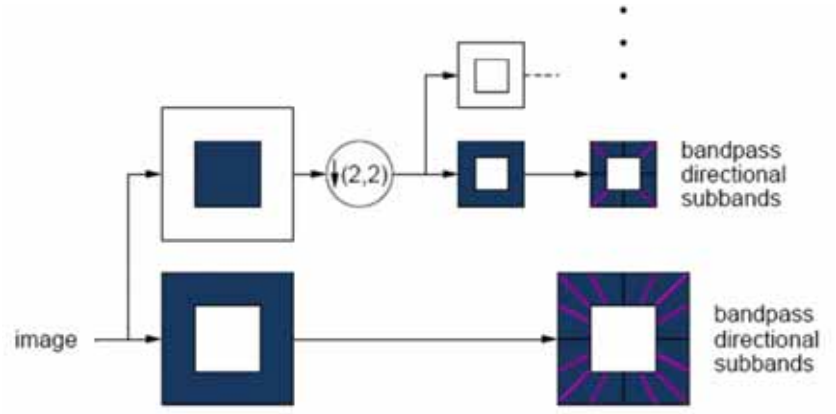

Fig. 6. The contourlet filterbank: first, a multiscale decomposition into octave bands by the Laplacian pyramid is computed, and then a directional filter bank is applied to each bandpass channel.

Although the contourlet Transform is easier to understand in the practical side, being a very elegant framework, the theoretical bases are not as robust as the ones in the curvelet Transform, in the sense that for most choices of filters in the angular filterbank, contourlets are not sharply localized in frequency, contrarily to the curvelet elements, whose location is sharply defined as the polar wedges of figure n. On the other hand, the contourlet transform is directly designed for discrete applications, whereas the discretization scheme of the curvelet transform faces some intrinsic challenges in the sampling of the Fourier plane in the outermost coronae, presenting the contourlet transform less redundancy also.
The potential of curvelet/contourlet based algorithms has been demonstrated in recent works. (Dettori & Semler 2007) compares the texture classification performance of wavelet, ridgelet and curvelet-based algorithms for CT tissue identification, where is evident that the curvelet outperforms the other methods. (Li & Meng 2009) states that the performance traditional texture extraction algorithms, in this case the local binary pattern texture operator, improves if applied in the curvelet domain. (Yang et al. 2008) proposed a contourlet-based image fusion scheme that presents better results than the ones achieved with wavelet techniques.

3. Basics on pattern recognition and hidden Markov models

3.1 Pattern recognition with HMM's

Hidden Markov Models (HMM's) make usually part of pattern recognition systems which basic principle applied to phonocardiography is shown in figure 7. An incoming pattern is classified according to a pre-trained dictionary of models. These models are in the present case HMM's, each one modeling each event in the phonocardiogram. The events are the four main waves M1, T1, A2 and P2, and the background that can accommodate systolic and diastolic murmurs. The pattern classification block evaluates the likelihood of A2 preceding P2 and vice versa and also the most likely state sequence for each hypothesis through the super HMM, which is constituted by the appropriate concatenation of the models in the

dictionary. The feature extraction block takes advantage of the WT to better discriminate the wave spectral content. The signal is simultaneously viewed at three different scales each one pointing out different signal characteristics.

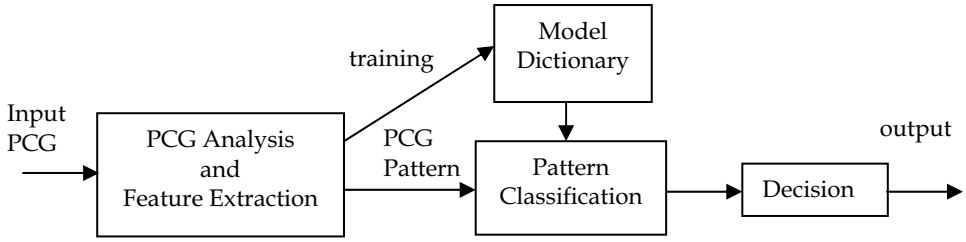

Fig. 7. Principle of a pattern recognition on PCG.

Such a system operates in two phases:

A **training phase**, during which the system learns the reference patterns representing the different PCG sounds (e.g. M1, T1, A2, P2 and background) that constitute the vocabulary of the application. Each reference is learned from labeled PCG examples and stored in the form of models that characterise the patterns properties. The learning phase necessitates efficient learning algorithms for providing the system with truly representative reference patterns.

A **recognition phase**, during which an unknown input pattern is identified by considering the set of references. The pattern classification is done computing a similarity measure between the input PCG and each reference pattern. This process necessitates defining a measure of closeness between feature vectors and a method for aligning two PCG patterns, which may differ in duration and cardiac rhythm.

By nature the PCG signal is neither deterministic nor stationary. Non-deterministic signals are frequently but not always modelled by statistical models in which one tries to characterise the statistical properties of the signal. The underlying assumption of the statistical model is that the signal can be characterised as a stochastic process, which parameters can be estimated in a precise manner. A stochastic model compatible with the non-stationary property is the Hidden Markov Model (HMM), which structure is shown in figure 4. This stochastic model consists of a set of states with transitions between them. Observation vectors are produced as the output of the Markov model according to the probabilistic transitioning from one state to another and the stationary stochastic model in each state. Therefore, the Markov model segments a non-stationary process in stationary parts providing a very rich mathematical structure for analysing non-stationary stochastic processes. So these models providing a statistical model of both the static properties of cardiac sounds and the dynamical changes that occur across them. Additionally these models, when applied properly, work very well in practice for several important applications besides the biomedical field.

3.2 Hidden Markov Models

Hidden Markov models are a doubly stochastic process in which the observed data are viewed as the result of having passed the hidden finite process (state sequence) through a function that produces the observed (second) process. The hidden process is a collection of states connected by transitions, each one described by two sets of probabilities:

A **transition probability**, which provides the probability of making a transition from one state to another.

An **output probability** density function, which defines the conditional probability of observing a set of cardiac sound features when a particular transition takes place. The continuous density function most frequently used is the multivariate Gaussian mixture.

In an HMM the goal of the decoding or recognition process is to determine a sequence of hidden (unobservables) states (or transitions) that the observed signal has gone through. The second goal is to define the likelihood of observing that particular event, given a state sequence determined in the first process. Given the Markov models definition, there are two problems of interest:

The **Evaluation Problem**: Given a model and a sequence of observations, what is the probability that the observations are generated by the model? This solution can be found using the forward-backward or Baum-Welch algorithm (Baum 1972, Rabiner 1989).

The **Learning Problem**: Given a model and a sequence of observations, what should the model's parameters be, so that it has the maximum likelihood of generating the observations? This solution can be found using the Baum-Welch or forward-backward algorithm (Baum 1972).

3.2.1 The evaluation problem

The goal of this and the next sub-section is not to broach exhaustively the HMMs theory, but only provide a basis to help in best understanding how these flexible stochastic models can be adapted to several modeling situations regarding biomedical applications. More details can be encountered in (Rabiner 1989).

When the random variables of a Markov Process take only discrete values, (frequently integers, the states are numerated by integer values) the stochastic state machine is known by Markov chain. If the state transition at each time is only dependent of the previous state, then the Markov chain is said of first order. The HMMs reviewed in this chapter are first order Markov chains.

Consider a left to right connected HMM with 6 states as illustrated in Figure 8 (for simplicity, the density probability function is not shown).

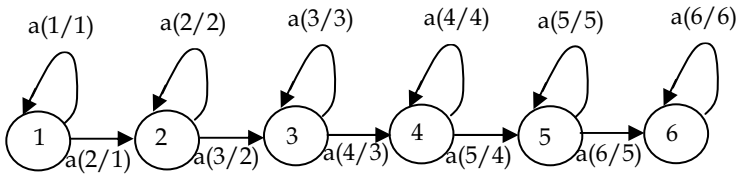

Fig. 8. A left to right HMM with 6 states

This stochastic state machine is characterised by the state transition matrix A, the probability density function in each transition B and the initial state probability vector π. The PCG signal is characterised by a time evaluating event sequence, whose properties change over time in a successive manner. Furthermore, as time increases, the state index increases or stays the same, that is, the system states proceed from left to right, and the state sequence must begin in state 1 and end in the last one for a cardiac cycle begining in an S1 sound. In this conditions a(i/j)=0, j>i and π_i have the property

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \tag{9}$$

As at each time the transition comes up then a(./i)=1, where a(./i) stands for transition from state i to each other. The transition dependent probability density function is typically a finite Gaussian multivariate mixture of the form

$$f(\mathbf{y}/s_t) = \sum_{c_t=1}^{C} p_{s_t,c_t} G\left(\mathbf{y}_t, \boldsymbol{\mu}_{s_t,c_t}, \boldsymbol{\Sigma}_{s_t,c_t}\right) \qquad 1 \le s_t \le N \tag{10}$$

where $\mathbf{y}$ is the observation vector being modelled, p_{s_t,c_t} is the mixture coefficient for the c[th] mixture in state s at time t, G(.) stands for Gaussian (Normal) distribution, and N is the number of states in the model. Other types of log-concave or elliptical distributions can be used (Levinson et al. 1983).

Given a sequence of vector observations $\mathbf{Y}=\{\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_T\}$, what is the likelihood that the model generated the observations? As an example suppose T=11, and the model shown in Figure 8. One possible time indexed path through the model is 1r, 1n, 2r, 2n, 3r, 3n, 4r, 4n, 5r, 5n, 6r, when r stands for recursive transitions and n stands for next transitions. Another possible path is 1r, 1r, 1r, 1n, 2n, 3n, 4n, 5n, 6r, 6r, 6r. As the model generates observations that can arrive from any path (events mutually exclusives) then the likelihood of the sequence is the sum of the likelihood in each path. Let $\mathbf{s}=\{s_1, s_2, \ldots s_T\}$ be one considered state sequence. The likelihood of the model generates the observed vector sequence $\mathbf{Y}$ given one such fixed-state sequence $\mathbf{S}$ and the model parameters $\lambda=\{A,B,\pi\}$ is given by

$$P(\mathbf{Y}/\mathbf{S},\lambda) = f(\mathbf{y}_1/s_1,\lambda).f(\mathbf{y}_2/s_2,\lambda)...f(\mathbf{y}_T/s_T,\lambda) = \prod_{t=1}^{T} f(\mathbf{y}_t/s_t,\lambda) \tag{11}$$

The probability of such a state sequence **S** can be written as

$$P(\mathbf{S}/\lambda) = \pi_{s_1} a_{s_1 s_2} a_{s_2 s_3} ... a_{s_{T-1} s_T} \tag{12}$$

The joint probability of **Y** and **S**, i.e., the probability that **Y** and **S** occur simultaneously, is simply the product of the above two terms

$$f(\mathbf{Y},\mathbf{S}/\lambda) = f(\mathbf{Y}/\mathbf{S},\lambda)P(\mathbf{S}/\lambda) \tag{13}$$

The probability of **Y** (given the model) is obtained by summing this joint probability over all possible state sequences **S** and is given by

$$f(\mathbf{Y}/\lambda) = \sum_{S} f(\mathbf{Y}/\mathbf{S},\lambda)P(\mathbf{S}/\lambda) = \sum_{s_1,s_2,...s_T} \pi_{s_1} f(\mathbf{y}_1/s_1,\lambda) a_{s_1 s_2} f(\mathbf{y}_2/s_2,\lambda) ... a_{s_{T-1} s_T} f(\mathbf{y}_T/s_T,\lambda) \tag{14}$$

The interpretation of the computation in the above equation is the following. Initialy (at time t=1) the HMM is in state s_1 with probability π_{s1}=1, and generates the symbol $\mathbf{y}_1$ (in this state/transition) with probability $f(\mathbf{y}_1/s_1,\lambda)$. The clock changes from time *t* to *t+1* (time=2) and the HMM make a transition to s_2 from state s_1 with probability a_{s1s2} and generates symbol $\mathbf{y}_2$ with probability $f(\mathbf{y}_{21}/s_2,\lambda)$. This process continues until the last transition (at time T) from state s_{T-1} to state s_T with probability $a_{sT-1\ sT}$ and generates symbol $\mathbf{y}_T$ with probability $f(\mathbf{y}_T/s_T,\lambda)$.

To conclude this section it is convenient rewrite the equation (14) in a more compact and useful form. Thus, substituting (10) in (14) we obtain

$$f(\mathbf{Y}/\lambda) = \sum_{S} \prod_{t=1}^{T} a_{s_{t-1} s_t} f(\mathbf{y}_t/s_t,\lambda) = \sum_{S} \prod_{t=1}^{T} a_{s_{t-1} s_t} \sum_{c_t=1}^{C} p_{s_t,c_t} G(\mathbf{y}_t,\boldsymbol{\mu}_{s_t,c_t},\boldsymbol{\Sigma}_{s_t,c_t}) \tag{15}$$

or in a more suitable and general form

$$f(\mathbf{Y}/\lambda) = \sum_{S} \sum_{C} \prod_{t=1}^{T} a_{s_{t-1} s_t} p_{s_t,c_t} f(\mathbf{y}_t/s_t,c_t,\lambda) \tag{16}$$

3.2.2 The evaluation problem

The most difficult problem of HMMs is to determine a method to adjust the model parameters (A,B,π) to satisfy a certain optimisation criterion. There is no known way to analytically solve for the model parameter set that maximises the probability of the observation sequence in a closed form. It can be, however, choose λ=(A,B,π) such that its likelihood P(**Y**/λ), is locally maximised using an iterative procedure such as the Baum-Welch method (also known as the Expectation Maximisation (EM) method) or using gradient techniques (Levinson et al. 1983). This sub-section shows the ideas behind the EM algorithm, showing it usefulness in the resolution of problems with missing data.

Hidden Markov models are a doubly stochastic process where the first, the state sequence, is unobserved and so unknown. The observed vector sequences (observable data) are called incomplete data because they are missing the unobservable data, and data composed by observable and unobservable data are called complete data. Making use of the observed (incomplete) data and the joint probability density function of observed and unobserved data, the EM algorithm iteratively maximises the log-likelihood of observable data.
In the particular HMM case, there are a measure space **S** (state sequence) of unobservable data, corresponding to a measure space **Y** (observations) of incomplete data. Here **Y** is easy to observe and measure, while **S** contains some hidden information that is unobservable. Let f(**s**/λ) and f(**y**/λ) be members of a parametric family of probability density functions (pdf) defined on **S** and **Y** respectively for parameter λ. For a given **y**ε**Y**, the goal of the EM algorithm is to maximise the log-likelihood of the observable data **y,** L(**y**,λ)=log f(**y**/λ), over λ by exploiting the relationship between f(**y**,**s**/λ) and f(**s**/**y**,λ). The joint pdf f(**y**,**s**/λ) is given by

$$f(\mathbf{y},\mathbf{s}/\lambda) = f(\mathbf{s}/\mathbf{y},\lambda) f(\mathbf{y}/\lambda) \tag{17}$$

From the above expression the following log-likelihood can be obtained

$$\log f(\mathbf{y}/\lambda) = \log f(\mathbf{y},\mathbf{s}/\lambda) - \log f(\mathbf{s}/\mathbf{y},\lambda) \tag{18}$$

and for two parameter sets λ' and λ, the expectation of incomplete log-likelihood L(**y**,λ′) over complete data (**y**,**s**) conditioned by **y** and λ is

$$\mathrm{E}_s\left[L(\mathbf{y},\lambda')/\mathbf{y},\lambda\right] = \mathrm{E}\left[\log f(\mathbf{y}/\lambda'/\mathbf{y},\lambda\right] = \int \log f(\mathbf{y}/\lambda') f(\mathbf{s}/\mathbf{y},\lambda) ds$$

$$= \log f(\mathbf{y}/\lambda') = L(\mathbf{y},\lambda') \tag{19}$$

where E[./**y**,λ] is the expectation conditioned by **y** and λ over complete data (**y**,**s**). Then from (18) the following expression is obtained

$$L(\mathbf{y},\lambda') = Q(\lambda,\lambda') - H(\lambda,\lambda') \tag{20}$$

where

$$Q(\lambda,\lambda') = \mathrm{E}_s\left[\log f(\mathbf{y},\mathbf{s}/\lambda')/\mathbf{y},\lambda\right] \tag{21}$$

and

$$\mathrm{H}(\lambda,\lambda') = \mathrm{E}_s\left[\log f(\mathbf{s}/\mathbf{y},\lambda')/\mathbf{y},\lambda\right] \tag{22}$$

The basis of the EM algorithm lies in the fact that if Q(λ,λ′)≥Q(λ,λ), then L(y,λ′)≥L(y,λ), since it follows from Jensen's inequality that H(λ,λ′)≤H(λ,λ) (Dempster et al. 1977). This fact implies that the incomplete log-likelihood L(y,λ) increases monotonically on any iteration of parameter update from λ to λ′, via maximisation of the Q function which is the expectation of log-likelihood from complete data.

From equation (15) and for the complete data we have

$$f(\mathbf{Y},\mathbf{S},\mathbf{C}/\lambda') = \prod_{t=1}^{T} a'_{s_{t-1}s_t}\, p'_{s_t,c_t}\, f(\mathbf{y}_t/s_t,c_t,\lambda') \tag{23}$$

and from equation (20) we obtain

$$Q(\lambda,\lambda') = \mathrm{E}\left[\log P(\mathbf{Y},\mathbf{S},\mathbf{C}/\lambda')/\mathbf{Y},\lambda\right] = \sum_{S}\sum_{C} P(\mathbf{Y},\mathbf{S},\mathbf{C}/\lambda)\log f(\mathbf{Y},\mathbf{S},\mathbf{C}/\lambda') \tag{24}$$

substituting equation (23) in (24) we obtain

$$\begin{aligned} Q(\lambda,\lambda') &= \sum_{S}\sum_{C} P(\mathbf{Y},\mathbf{S},\mathbf{C}/\lambda)\log\prod_{t=1}^{T} a'_{s_{t-1}s_t}\, p'_{s_t,c_t}\, f(\mathbf{y}_t/s_t,c_t,\lambda') \\ &= \sum_{S}\sum_{C} P(\mathbf{Y},\mathbf{S},\mathbf{C}/\lambda)\sum_{t=1}^{T}\left\{\log a'_{s_{t-1}s_t} + \log p'_{s_t,c_t} + \log f(\mathbf{y}_t/s_t,c_t,\lambda')\right\} \end{aligned} \tag{25}$$

At this point it is finished the expectation step of the EM algorithm. Equation (24) shows that the Q function is separately in three independent terms, one is state transition dependent, another is component mixture dependent and the last is dependent of the pdf parameters of observation incomplete data. In the second step of the EM algorithm known as the maximisation step, the Q function is maximised in order to the parameters to be estimated. For example to estimate the matrix A, the Q function must be maximised in order to the respective parameters under the constraint

$$\sum_{j=1}^{N} a'(j/i) = 1 \tag{26}$$

i.e. at each time clock the transition must occur. To estimate the mixture coefficients, the probability over all the space must be one, and express as the following constraint:

$$\sum_{c_t=1}^{C} p'_{i,c_t} = 1 \qquad 1 \le i \le N \tag{27}$$

Understanding the fundamental concepts of the EM algorithm the derivation of the reestimation formulas is straightforward. First of all we can address the most general case where the initial state is not known and must be estimated. In this situation the auxiliary Q function can be written from equations (12), (13), (14) and (25) as

$$Q(\lambda,\lambda') = \sum_{\mathbf{S}}\sum_{\mathbf{C}} P(\mathbf{Y},\mathbf{S},\mathbf{C}/\lambda)\left\{\log \pi'_{s_1} + \sum_{t=1}^{T-1}\log a'_{s_t,s_{t+1}} + \sum_{t=1}^{T}\log p'_{s_t,c_t} + \sum_{t=1}^{T}\log f(\mathbf{y}_t/s_t,c_t,\lambda')\right\} \tag{28}$$

The auxiliary Q function can be maximized separately in order to each term, so regarding to the initial state vector the Q function can be written as

$$Q_\pi(\lambda,\pi') = \sum_{\mathbf{S}}\sum_{\mathbf{C}} f(\mathbf{Y},\mathbf{S},\mathbf{C}/\lambda)\log \pi'_{s_1} = \sum_{j}\sum_{\mathbf{C}} f(\mathbf{Y},s_1=j,c_t/\lambda)\log \pi'_{s_j} \tag{29}$$

which results in an equation of the type

$$\sum_{j=1}^{N} w_j \log y_j \quad \text{under the constraint} \quad \sum_{j=1}^{N} y_j = 1 \tag{30}$$

Equation (29) has a global maximum at

$$y_j = \frac{w_j}{\sum_{i=1}^{N} w_i} \qquad j=1,2,...,N \tag{31}$$

Using equation (31) in the solution of equation (29) we obtain

$$\pi'_{s_j} = \frac{\sum_{\mathbf{C}} f(\mathbf{Y},s_1=j,c_t/\lambda)}{\sum_{j}\sum_{\mathbf{C}} f(\mathbf{Y},s_1=j,c_t/\lambda)} = \frac{f(\mathbf{Y},s_1=j/\lambda)}{\sum_{j} f(\mathbf{Y},s_1=j/\lambda)} = \frac{f(\mathbf{Y},s_1=j/\lambda)}{f(\mathbf{Y}/\lambda)} \tag{32}$$

Similarly the part of the auxiliary Q function regarding to the state transition matrix can be written as

$$Q(\lambda,a'_{i,j}) = \sum_{\mathbf{S}}\sum_{\mathbf{C}} f(\mathbf{Y},\mathbf{S},\mathbf{C}/\lambda)\sum_{t=1}^{T-1}\log a'_{i,j} = \sum_{i} Q_{a_i}(\lambda,a'_{i,j}) \tag{33}$$

For a particular state *i* the sum in **S** in the second member of equation (32) disapears. However, as for each state i the probability of transition for any state *j* is the sum of the transition probabilities to all possible states *j* (including the state *i* itself) the individual Q function regarding the state transition probabilities, for a given state *i* can be written from equation (32) as

$$Q_{a_i}(\lambda,a'_{i,j}) = \sum_{j}\sum_{t=1}^{T-1}\sum_{\mathbf{C}} f(\mathbf{Y},s_t=i,s_{t+1}=j,c_t/\lambda)\log a'_{i,j} \tag{34}$$

From equation (31) the maximization of equation (34) can be written as

$$a'_{i,j} = \frac{\sum_{t=1}^{T-1}\sum_{\mathbf{C}} f(\mathbf{Y}, s_t = i, s_{t+1} = j, c_t / \lambda)}{\sum_{j}\sum_{t=1}^{T-1}\sum_{\mathbf{C}} f(\mathbf{Y}, s_t = i, s_{t+1} = j, c_t / \lambda)} = \frac{\sum_{t=1}^{T-1} f(\mathbf{Y}, s_t = i, s_{t+1} = j / \lambda)}{\sum_{t=1}^{T-1} f(\mathbf{Y}, s_t = i / \lambda)} \tag{35}$$

Regarding the mixture coefficients, the individual Q function can be written from equation (28) as

$$Q(\lambda, p'_{j,c}) = \sum_{\mathbf{S}}\sum_{\mathbf{C}} f(\mathbf{Y}, \mathbf{S}, \mathbf{C} / \lambda) \sum_{t=1}^{T} \log p'_{j,c} = \sum_{j} Q_{p_j}(\lambda, p'_{j,c}) \tag{36}$$

For a particular state *j* equation (36) can be written as

$$Q_{p_j}(\lambda, p'_{j,c}) = \sum_{\mathbf{C}} f(\mathbf{Y}, \mathbf{S}, \mathbf{C} / \lambda) \sum_{t=1}^{T} \log p'_{j,c} = \sum_{c=1}^{C}\sum_{t=1}^{T} f(\mathbf{Y}, s_t = j, c_t = c / \lambda) \log p'_{j,c} \tag{37}$$

Which solution, obtained from equation (30) is

$$p'_{j,c} = \frac{\sum_{t=1}^{T} f(\mathbf{Y}, s_t = j, c_t = c / \lambda)}{\sum_{c=1}^{C}\sum_{t=1}^{T} f(\mathbf{Y}, s_t = j, c_t = c / \lambda)} = \frac{\sum_{t=1}^{T} f(\mathbf{Y}, s_t = j, c_t = c / \lambda)}{\sum_{t=1}^{T} f(\mathbf{Y}, s_t = j / \lambda)} \tag{38}$$

Regarding the distribution parameters (excluding the mixture coefficients) the Q function is

$$Q(\lambda, f'_{s_t, c_t}) = \sum_{\mathbf{S}}\sum_{\mathbf{C}} f(\mathbf{Y}, \mathbf{S}, \mathbf{C} / \lambda) \sum_{t=1}^{T} \log f(\mathbf{y}_t / s_t, c_t, \lambda')$$

$$= \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{c=1}^{C} f(\mathbf{y}_t / s_t, c_t, \lambda) \log f(\mathbf{y}_t / s_t, c_t, \lambda') = \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{c=1}^{C} \gamma_t(n,c) \log f(\mathbf{y}_t / s_t, c_t, \lambda') \tag{39}$$

Where $\gamma_t(n,c)$ is the joint probability density function of the observation vector $\mathbf{y}_t$, the state n and the mixture component c. Assuming the observations independents and identically distributed (iid) and with Gaussian distribution, equation (39) can be written as

$$Q(\lambda, f'_{s_t,c_t}) = \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{c=1}^{C}\gamma_t(n,c)\log\prod_{i=1}^{D}G(y_{t,i}, \mu'_{n,c,i}, \sigma'^2_{n,c,i}) \tag{40}$$

Where $y_{t,i}$ is the ith component of the observation vector at time t, $\mu_{n,c,i}$ and $\sigma^2_{n,c,i}$ are respectively the mean and variance of the ith component of mixture c in state n and D is the dimensionality of the observation vector. Substituting the Gaussian function in equation (40) we obtain

$$Q(\lambda, f'_{s_t,c_t}) = -\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{c=1}^{C}\gamma_t(n,c)\sum_{i=1}^{D}\left[\frac{1}{2}\log\sigma'^2_{n,c,i} + \frac{(y_{t,i} - \mu'_{n,c,i})^2}{2\sigma'^2_{n,c,i}}\right] \tag{41}$$

The solution for the maximization of equation (41) is in general obtained by differentiation. For the mean we have

$$\frac{dQ(\lambda, f'_{s_t,c_t})}{d\mu'_{n,c,i}} = \sum_{t=1}^{T}\gamma_t(n,c)\frac{2}{2\sigma'^2_{n,c,i}}(y_{t,i} - \mu'_{n,c,i}) = 0 \tag{42}$$

Which solution is

$$\mu'_{n,c,i} = \frac{\sum_{t=1}^{T}\gamma_t(n,c)y_{t,i}}{\sum_{t=1}^{T}\gamma_t(n,c)} \tag{43}$$

Differentiating equation (41) in order to variance we obtain

$$\frac{dQ(\lambda, f'_{s_t,c_t})}{d\sigma'^2_{n,c,i}} = -\sum_{t=1}^{T}\gamma_t(n,c)\left\{\frac{1}{2\sigma'^2_{n,c,i}} - \frac{(y_{t,i} - \mu'_{n,c,i})^2}{4\sigma'^4_{n,c,i}}\right\} = 0 \tag{44}$$

Which solution is given by

$$\sigma'^2_{n,c,i} = \frac{\sum_{t=1}^{T}\gamma_t(n,c)(y_{t,i} - \mu'_{n,c,i})^2}{\sum_{t=1}^{T}\gamma_t(n,c)} \tag{45}$$

The reestimation formulas given by equations (45), (43), (38), (35) and (32) can be easily calculated using the definitions of forward sequence $\alpha_t(i)=f(\mathbf{y}_1,\mathbf{y}_2,...\mathbf{y}_t,s_t=i/\lambda)$ and backward

sequence $\beta_t(i)=f(\mathbf{y}_{t+1},\mathbf{y}_{t+2},...\mathbf{y}_T,s_t=i/\lambda)$. This procedure is standard in the HMM implementation.

4. Wavelets, HMM's and Bioacoustics

Recently a new approach based on wavelets and HMM's was suggested for PCG segmentation purposes (Lima & Barbosa 2008). The main idea is to take advantage of the ability of HMM's to break non-stationary signals in stationary segments modelling both the static properties of cardiac sounds and the dynamical changes that occur across them. However the cardiac sound is particularly difficult to analyse since some events that must be identified are of very close characteristics, and are frequently corrupted by murmurs which are noise-like events very important concerning the diagnosis of several pathologies such as valvular stenosis and insufficiency. This approach takes also advantage of the WT to emphasize the small differences between similar events viewed at different scales, while the scales less affected by noise can be chosen for analysis purposes.
A normal cardiac cycle contains two major sounds: the first heart sound S1 and the second heart sound S2. S1 occurs at the onset of ventricular contraction and corresponds in timing to the QRS complex. S2 follows the systolic pause and is caused by the closure of the semilunar valves. The importance of S2 regarding diagnosis purposes has been recognized for a long time, and its significance is considered of utmost importance, by cardiologists, to auscultation of the heart (Leatham 1987). This approach concentrates mainly on the analysis of the second heart sound (S2) and its two major components A2 and P2. The main purposes are estimating the order of occurrence of A2 and P2 as well as the time delay between them. This delay known as *split* occurs from the fact that the aortic and pulmonary valves do not close simultaneously. Normally the aortic valves close before the pulmonary valves and exaggerated splitting of the S2 sound may occur in right ventricular outflow obstruction, such as pulmonary stenosis (PS), right bundle branch block (RBBB) and atrial and ventricular septal defect. Reverse splitting of sound S2 is due to a delay in the aortic component A2, which causes a reverse sequence of the closure sounds, with P2 preceding A2. The main causes of reverse spitting are left bundle branch block (LBBB) and premature closure of pulmonary valves. The wide *split* has duration of about 50 miliseconds compared to the normal *split* with the value of ≤ 30 ms (Leung et al. 1998). The measurement of the S2 *split*, lower or higher than 30 ms and the order of occurrence of A2 and P2 leads to a discrimination between normal and pathological cases.

4.1 Wavelet Based feature extraction

The major difficulty associated with the phonocardiogram segmentation is the similarity among its main components. For example it is well known that S1 and S2 contain very closed frequency components, however S2 have higher frequency content than S1. Another example of sounds containing very closed frequency components, which must be distinguished is the aortic and pulmonary components of S2 sound.

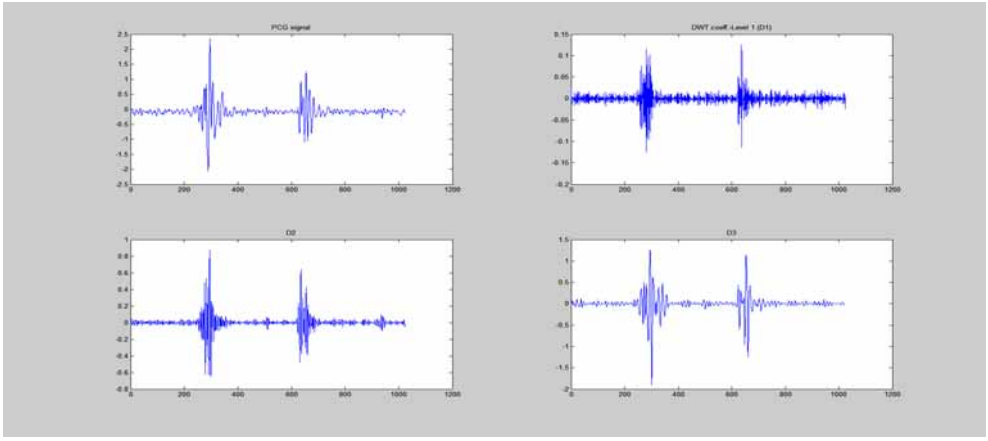

Fig. 9. Wavelet decomposition of one cycle PCG

The multiresolution analysis based on the DWT can enhance each one of these small differences if the signal is viewed at the most appropriate scale. Figure 9 shows the result of the application of the DWT one cycle of a normal PCG. From the figure we can observe that d1 level (frequency ranges of 250-500Hz) emphasize the high frequency content of S2 sound when compared with S1. D2 and d3 levels show clearly the differences in magnitude and frequency of the S2 components A2 and P2, which helps to accurately measure the *split* since A2 and P2 appear quite different. The features used in the scope of this work are simultaneous observations of d1, d3 and d4 scales, therefore the observation sequence generated after the parameter extraction is of the form $\mathbf{O}=(\mathbf{o}_1, \mathbf{o}_2, \ldots \mathbf{o}_T)$ where T is the signal length in number of samples and each observation $\mathbf{o}_t$ is a three-dimensional vector, i. e., the wavelet scales have the same time resolution as the original signal.

4.2 HMM segmentation of the PCG

The phonocardiogram can be seen as a sequence of elementary waves and silences containing at least five different segments; M1, T1, A2, P2 and silences. Each one of them can be modeled by a different HMM. Two different silences must be modeled since murmurs can be present and diastolic murmurs are in general different from systolic murmurs. Left to right (or Bakis) HMM's with different number of states were used, since this is the most used topology in the field of speech recognition, and the phonocardiographic signal is also a sound signal with auditory properties similar to speech signals. Each HMM models one different event and the concatenation of them models the whole PCG signal. The concatenation of these HMM's follows certain rules dependent on the sequence of events allowed. These rules define a grammar with six main symbols (four main waves and two silences of different nature) and an associated language model as shown in figure 10.

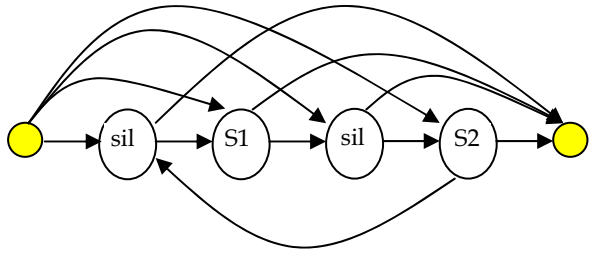

Fig. 10. Heart sound Markov Model

This HMM does not take into consideration the S3 and S4 heart sounds since these sounds are difficult to hear and record, thus they are most likely not noticeable in the records.
The acoustic HMM's are Continuous Density Hidden Markov Models (CDHMM's) and the continuous observations are frequently modeled by a Gaussian mixture. However, by observing the histograms for every state of every HMM it was observed that most of them appear to be well fitted to a single Gaussian, so a single Gaussian probability density function was used to model the continuous observations in each state/transition.
PCG elementary waves are modeled by three state HMM's and the probability density functions are simple Gaussians. The observation vector components are considered independents and identically distributed as considered in the re-estimation formulas in section 3. Silence models are one state HMM's and probabilities density functions are a mixture of three Gaussian functions.
The PCG morphologies are learned from training the HMM's. The training algorithm was the standard Baum-Welch method, also called forward-backward algorithm, which is a particular case of the expectation maximization method and is extensively explained in section 3.
The beat segmentation procedure consists on matching the HMM models to the PCG waveform patterns. This is typically performed by the Viterbi decoding algorithm, which relates each observation to an HMM state following the maximum likelihood criteria with respect to the beat model structure. Additionally the most likely state sequence is available, which allows to estimate time duration of the PCG components as the *split*. This algorithm performs well in the absence of strong murmurs. However if relatively strong murmurs are present both the silence models must be adapted for the current patient, even if murmurs exist in the training patterns. Two methods are suggested:
If the ECG is also recorded a QRS detector can be used to accurately locate diastolic murmurs that appear exactly before QRS locations. Systolic murmurs locations can also be estimated since they appear after S1 that is almost synchronous with the QRS. Having systolic and diastolic data the corresponding silence models can be updated for the current patient by using incremental training. Three cardiac cycles are enough to accurately re-estimate the silence models. Additionally using the re-estimated silence models all the other models can be updated for the current patient by using also incremental training or adaptation. Firstly the most likely wave sequence is estimated by decoding the data, then all the models except the silence models are updated on the basis of the recognition data. Two cardiac cycles are enough to adapt the wave models. This procedure incremented the system

performance of 17.25% when applied to a patient with systolic murmur, suspection of pulmonary stenosis, ventricular septal defect and pulmonary hypertension.
In the absence of ECG the most likely wave sequence can also be estimated by decoding the data and all models can be updated based on incoming data by using the formulas derived in section 3. However under severe murmur conditions the decoding can fail and the updating of the models originates model divergence. Therefore supervised adaptation is required to guarantee model convergence. Under model convergence situations and using two cardiac cycles for model adaptation purposes similar results to the previous case were obtained in the same dataset.
The performance of this algorithm is similar to the performance of (Debbal & Bereksi-Reguig 2007) algorithm in the absence of murmurs and in the most common situation where the aortic wave has higher amplitude than the pulmonary wave. However in the presence of a relatively weak systolic murmur in real data as well in noisy situations the present algorithm outperformed the (Debbal & Bereksi-Reguig 2007) algorithm.

5. Wavelets, HMM's and the ECG

Recently WT has been successfully combined with HMM's providing reliable beat segmentation results (Andreão et al., 2006). The ECG signal is decomposed in different scales by using the DWT and re-synthesized using only the most appropriate scales. Three views of the ECG at different scales were used in such a way that the re-synthesized signal has the same time resolution as the original ECG. Each wave (P, QRS, T) and segment (PQ, ST) of a heartbeat is modelled by a specific left-to-right HMM. The isolelectric line between two consecutive beats is also modelled by an HMM. The concatenation of the individual HMM's models the whole ECG signal. The continuous observations are modelled by a single Gaussian probability density function, since histograms of the observations in the various HMM's showed that the data can be well fitted to a single Gaussian. In order to improves modelling of complex patterns multiple models are created for each waveform by using the HMM likelihood clustering algorithm.
A morphological based strategy in the HMM framework have recently been proposed to take advantage of the similarities between normal and atrial fibrillation beats to improve the classifier performance by using Maximum Mutual Information (MMIE) training, in a single model/double class framework (Lima & Cardoso 2007). The approach is similar of having two different models sharing the most parameters. This approach saves computational resources at run-time decoding and improves the classification accuracy of very similar classes by using MMIE training. The idea is that if two classes have some state sequence similarities and the main morphological differences occur only in a short time slice, then setting appropriately internal state model transitions can model the differences between classes. These differences can be more efficiently emphasized by taking advantage of the well known property of MMIE training of HMM's, which typically makes more effective use of a small number of available parameters. By this reasoning the selected decoding class can be chosen on the basis of the most likely state sequence, which characterizes the most likely class.
Figure 11 shows the model structure for the atrial fibrillation and normal beats, where $a_{i,j}$ stands for transition probability from state i to state j. The behind reasoning is based on the assumption that an AF beat is similar to a normal beat without the P wave which can be

modeled by a transition probability that not pass through the state which models the P wave. The recursive transition in each state can model rhythm differences by time warping capabilities. At the end of the decoding stage the recognized class can be selected by searching (backtracking) the most likely state sequence. This structure can be seen as two separate HMM's sharing the most parameters. This parameter sharing procedure is justified by the fact that ventricular conduction is normal in morphology for AF beats, and we intend to use a limited amount of parameters, just the pdf associated with the transitions from state 5 to states 6 and 7, state 6 to itself and to state 7 to reinforce the discriminative power between classes. The separation between these two classes can be increased by using an efficient discriminative training as MMIE obtained on the basis of the parameters associated with the intra-class differences, just those above mentioned. It is very important to note that this approach reinforces the HMM distance among different model structures while the distance of HMM's in the same structure (those that share parameters) are obviously decreased. However, it is believed that an appropriate discriminative training can efficiently separate the classes modeled by the same HMM. Although a recognition system fully trained by using the MMIE approach can be more effective it surely needs a much degree of computational requirements in both training and run time decoding.

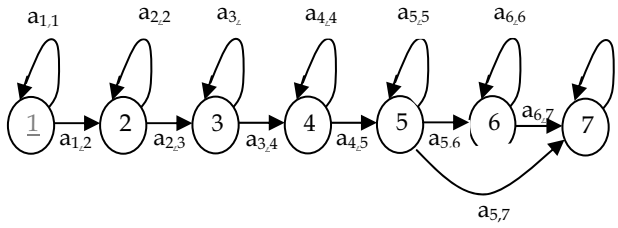

Fig. 11. HMM topology adopted for modelling normal (N) and atrial fibrillation (AF) beats.

States from 1 to 7 are concerned to the ECG events R, S, S-T, T, T-P, P, P-R.This frame state allocation concerned to the ECG events can be forced by setting (to one) the initial probability of the first state in the initial state probability vector and resetting all the other initial state probabilities, and also synchronizing the ECG feature extraction to begin in the R wave. This kind of synchronization is needed for this HMM topology where the initial state must be synchronized with the R wave, otherwise the assumption that state 6 models P-wave can not be true. We observed this evidence in our experiments. However if a back transition from the last to the initial state is added this synchronization is necessary only for the first ECG pulse decoding. The synchronization between ECG beats and the HMM model is facilitated by the intrinsic difference between the last and first state, since the last state models an isoelectric segment (weak signal) while the first state models the R wave which is a much strong signal. In other words if the HMM is in state 7 modeling an isoelectric segment the happening of a strong R wave tends to force a transition to state one which helps in model/beat synchronization. The adopted training strategy accommodate both the MMIE training and parameter sharing, or in other words an MMIE training procedure in only one HMM platform with capabilities to model two classes must be required. This compromise was obtained by estimating the shared parameters in the MLE sense. This

algorithm was tested in the MIT_BIH arrhythmia database and outperforms the traditional MLE estimation algorithm.

6. Conclusion

This chapter provides a review of the WT and points out its most important properties regarding non-stationary biosignal modelling, including the extension to biomedical image processing. However practical situations often require high accurate methods capable of handling, usually by training, highly non-stationary conditions. To cope with this variability a new PCG segmentation approach was proposed relying on knowledge acquired from training examples and stored in statistical quasi-stationary models (HMM's) with features obtained from the wavelet transform. The proposed algorithm outperforms a recent wavelet only based algorithm especially under relatively light murmur situations, which are the most common in practical situations. Additionally a recent HMM algorithm based on morphological concepts concerning to arrhythmia classification was reviewed. This approach is also new and outperforms the conventional HMM training strategies.

7. References

Addison, P. S., Watson, J. N., Clegg, J. R., Holzer, M., Sterz, F. & Robertson, C. E. (2000). Evaluating arrhythmias in ECG signals using wavelet transforms. *IEEE Eng. Med. Biol.,* Vol. 19, page numbers (104-109).

Akay, Y. M., Akay, M., Welkovitz, W., & Kostis, J. (1994). Noninvasive detection of coronary artery disease. *IEEE Eng. In Med. And Biol. Mag.*, vol. 13 nº5, page numbers (761-764).

Akay, M., & Szeto, H. H., (1994). Wavelet analisis of opioid drug effects on the elctrocortical activity in fetus. *Proc. Conf. Artif. Neural Networks in Eng.*, page numbers (553-558).

Al-Fahoum, A. S. & Howitt, I. (1999). Combined wavelet transformation and radial basis neural network for classifying life-threatening cardiac arrhythmias. *Med. Biol. Eng. Comput.,* Vol. 37, page numbers (566-573).

Andreão, R. V., Dorizzi, B. & Boudy, J. (2006). ECG analysis using hidden Markov models. *IEEE Transactions on Biomedical Engineering,* Vol. 53, No. 8, page numbers (1541-1549).

Barbosa, D., Ramos, J., Tavares, A. & Lima, C. S. (2009). Detection of Small Bowel Tumors in Endoscopic Capsule Images by Modeling Non-Gaussianity of Texture Descriptors. *International Journal of Tomography & Statistics, Special Issue on Image Processing* ISSN 0972-9976. In press.

Baum, L. (1972). An inequality and associated maximisation technique in statistical estimation of probabilistic functions of Markov processes. *Inequalities,* Vol. 3, page numbers (1-8).

Benedetto, J. J., & Teolis, A. (1993). A wavelet auditory model and data compression. Appl. Computat. Harmonic. Anal., Vol. 1, page numbers (3-28).

Candès, E. & Donoho, D. (2000). Curvelets - a surprisingly effective nonadaptive representation for objects with edges. *Curves and Surfaces*, L. L. Schumaker et al., (Ed.), page numbers (105-120), Vanderbilt University Press, Nashville, TN

Candès, E.; Demanet, L.; Donoho, D. & Ying, L. (2006). Fast discrete curvelet transforms, *SIAM Multiscale Modeling Simul,* Vol. 5, No.3, September 2006, page numbers (861-899)

Chebil, J. & Al-Nabulsi, J. (2007). Classification of heart sound signals using discrete wavelet Analysis. *International Journal of Soft Compting,* Vol. 2, No. 1, page numbers (37-41).

Daugman, J. G. (1988). Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoust Speech and Signal Process.,* Vol. 36, (July. 1988) page numbers (1169-1179).

Daugman, J. G. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Trans. Biomed. Eng.,* Vol. 36, (Jan. 1989) page numbers (107-114).

Debbal, S. M., & Bereksi-Reguig, F. (2004). Analysis of the second heart sound using continuous wavelet transform. *J. Med. Eng. Technol*., vol. 28, Nº 4, page numbers (151-156).

Debbal, S. M., & Bereksi-Reguig, F. (2007). Automatic measure of the split in the second cardiac sound by using the wavelet transform technique*, Computers in Biology and Medicine*, vol 37, page numbers (269-276).

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, Vol. 39, Nº1, page numbers (1-38).

Dettori L. & Semler, L. (2007). A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography, *Computers in Biology and Medicine,* Vol. 37, No. 4, April 2007, page numbers (486-498)

Dickhaus, H., Khadra, L., & Brachmann, J., (1994). Time-frequency analysis of ventricular late potentials, *Methods of Inform. in Med*., vol. 33 (2), page numbers (187-195).

Do, M. & Vetterli M. (2005). The Contourlet Transform: An Efficient Directional Multiresolution Image Representation, *IEEE Trans. on Image Processing*, Vol. 14, No. 12, December 2005, page numbers (2091-2106)

Donoho, D. (1995). De-noising by Soft-thresholding, *IEEE Trans. Information Theory*, Vol. 41, No. 3, May 1995, page numbers (613-617)

Duverney, D., Gaspoz, J. M., Pichot, V., Roche, F., Brion, R., Antoniadis, A. & Barthelemy, J-C. (2002). High accuracy of automatic detection of atrial fibrillation using wavelet transform of heart rate intervals. *PACE,* Vol. 25, page numbers (457-462).

Gaudart, L., Crebassa, J. & Petrakian, J. P. (1993). Wavelet transform in human visual channels. *Applied Optics,* Vol. 32, No. 22, page numbers (4119-4127).

Govindan, A., Deng, G. & Power, J. (1997). Electrogram analysis during atrial fibrillation using wavelet and neural network techniques, *Proc. SPIE 3169,* pp. 557-562.

Heinlein, P.; Drexl, J. & Schneider, W. (2003). Integrated wavelets for enhancement of microcalcifications in digital mammography, *IEEE Trans. Med. Imag*., Vol. 22, March 2003, page numbers(402-413).

Hubel, D. H. (1982). Exploration of the primary visual cortex: 1955-1978. *Nature,* Vol. 299, page numbers (515-524).

Inoue, H. & Miyasaki, A. (1998). A noise reduction method for ECG signals using the dyadic wavelet transform. *IEICE Trans. Fundam.,* Vol. E81A, page numbers (1001-1007).

Jin, Y.; Angelini, E.; Esser, P. & Laine, A. (2003). De-noising SPECT/PET Images Using Cross-scale Regularization, *Proceedings of the Sixth International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI 2003)*, pp. 32-40, Montreal, Canada, November 2003.

Jin, Y.; Angelini, E. & Laine, A. (2004) Wavelets in Medical Image Processing: Denoising, Segmentation, and Registration, In: Handbook of Medical Image Analysis: Advanced Segmentation and Registration Models, Suri, J.; Wilson, D. & Laximinarayan, S., (Ed.), page numbers (305-358), Kluwer Academic Publishers, New York.

Kadambe, S., Murray, R. & Boudreaux-Bartels, G. F. (1999). Wavelet transform-based QRS complex detector. *IEEE Trans. Biomed. Eng.,* Vol. 46, page numbers (838-848).

Kalayci, T. & Ozdamar, O., (1995). Wavelet pre-processing for automated neural network detection of spikes. *IEEE Eng. in Med. and Biol. Mag.*, vol. 14 (2), page numbers (160-166).

Karkanis, A.; Iakovidis, D.; Maroulis, D.; Karras, D. & Tzivras, M. (2003). Computer-aided tumor detection in endoscopic video using color wavelet features, *IEEE Trans. Info. Tech. in Biomedicine,* Vol. 7, No. 3, September 2003, page numbers (142-152)

Khadra, L., Matalgah, M., El-Asir, B., & Mawagdeh, S. (1991). The wavelet transform and its applications to phonocardiogram signal analysis, In: *Med. Informat.*, vol. 16, page numbers (271-277).

Khadra, L., Dickhaus, H., & Lipp, A. (1993). Representations of ECG-late potentials in the time-frequency plane, In: J. *Med. Eng. and Technol.*, vol. 17 (6) page numbers (228-231).

Leatham, A. (1987). Auscultation and Phonocardiography: a personal view of the past 40 years. *Heart J., Vol.* 57 (B2).

Leman, H. & Marque, C. (200). Rejection of the maternal electrocardiogram in the electrohysterogram signal. *IEEE Trans. Biomed. Eng.,* Vol. 47, page numbers (1010-1017).

Lemaur, G.; Drouiche, K. & DeConinck, J. (2003). Highly regular wavelets for the detection of clustered microcalcifications in mammograms, *IEEE Trans. Med. Imag.*, Vol. 22, March 2003, page numbers (393-401)

Leung, T. S., White, P. R., Cook, J., Collis, W. B., Brown, E. & Salmon, A. P. (1998). Analysis of the second heart sound for diagnosis of paediatric heart disease. *IEE Proceedings - Science, Measurement and Technology,* Vol. 145, Issue 6, (November of 1998) page numbers (285-290).

Levinson, S. E., Rabiner, L. R. & Sondhi, M. M. (1983). An introduction to the application of the theory of probabilistic function of a Markov process to automatic speech recognition. *Bell System Tech. J.*, Vol. 62, N^{a} 4, page numbers (1035-1074).

Li, B. & Meng, Q. (2009). Texture analysis for ulcer detection in capsule endoscopy images, *Image and Vision Computing*, In Press

Li, C., & Zheng, C., (1993). QRS detection by wavelet transform, In: *Proc. Annu. Confl. on Eng. in Med. And Biol.*, vol. 15, page numbers (330-331).

Li, C., Zheng, C., Tai, C. (1995). Detection of ECG characteristic points using wavelet transforms. *IEEE Trans. Biomed. Eng.,* Vol. 42, page numbers (21-28).

Lima, C.S. & Cardoso, M. J. (2007). Cardiac Arrhythmia Detection by Parameters Sharing and MMI Training of Hidden Markov Models. *The 29th IEEE EMBS Annual International Conference EMBC07*, Lyon, France, 2007.

Lima, C. S. & Barbosa, D. (2008). Automatic Segmentation of the Second Cardiac Sound by Using Wavelets and Hidden Markov Models, *The 30th IEEE EMBS Annual International Conference EMBC08*, Vancouver, Canada, 2008.

Lima, C. S., Barbosa, D., Tavares, A., Ramos, J., Monteiro, L., Carvalho, L. (2008). Classification of Endoscopic Capsule Images by Using Color Wavelet Features, Higher Order Statistics and Radial Basis Functions, *The 30th IEEE EMBS Annual International Conference EMBC08*, Vancouver, Canada.

Mallat, S. G., (1989). Multifrequency channel decompositions of images and wavelet models, *IEEE Trans. Acoust., Speech and Signal Process. Patt.*, vol. 37, (December 1989) page numbers (2091-2110).

Mallat, S., & Zhong, S., (1992). Characterization of signals from multiscale edges, In: *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 14, page numbers (710-732).

Mallat, S., (1998). *A wavelet tour of signal processing*, Academic Press.

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Amer.* , Vol. 70, No. 11, page numbers (1297-1300).

Martinez, J. P., Almeida, R., Olmos, S., Rocha, A. P. & Laguna, P. (2004). A wavelet based ECG delineator: evaluation on standard data bases. *IEEE Trans. Biomed. Eng.,* Vol. 51, page numbers (570-581).

Nikoliaev, N., Gotchev, A., Egiazarian, K. & Nikolov, Z. (2001). Supression of electromyogram interference on the electrocardiogram by transform domain denoising. *Med. Biol. Eng. Comput.,* Vol. 39, page numbers (649-655).

Nigam, V. & Priemer, R. (2006). A Procedure to extract the Aortic and the Pulmonary Sounds from the Phonocardiogram, *Proceedings of the 28th Annual International Conference of the IEEE in Engineering in Medicine and Biology Society*, pp. 5715-5718, August 2006.

Obaidat, M. S., (1993). Phonocardiogram signal analysis: techniques and performance. *J. Med. Eng. and Technol.*, vol. 17, page numbers (221-227).

Papadopoulos, A.; Fotiadis, D. & Costaridou, L. (2008). Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques, *Computers in Biology and Medicine,* Vol. 38, No. 10, October 2008, page numbers (1045-1055)

Park, K. L., Lee, K. J. & Yoon H. R. (1998). Application of a wavelet adaptive filter to minimise distortion of the ST-segment. *Med. Biol. Eng. Comput.,* Vol. 36, page numbers (581-586).

Park, K. L., Khil, M. J., Lee, B. C., Jeong, K. S., Lee, K. J. & Yoon H. R. (2001). Design of a wavelet interpolation filter for enhancement of the ST-segment. *Med. Biol. Eng. Comput.,* Vol. 39, page numbers (1-6)

Porat, M. & Zeevi, Y. Y. (1989). Localised texture processing in vision: analysis and synthesis in Gaborian Space. *IEEE Trans. Biomed. Eng.,* Vol. 36, (Jan. 1989) page numbers (115-129).

Przelaskowski, A.; Sklinda, K.; Bargieł, P.; Walecki, J.; Biesiadko-Matuszewska, M. & Kazubek, M. (2007). stroke detection: Wavelet-based perception enhancement of computerized tomography exams, *Computers in Biology and Medicine,* Vol. 37, No. 4, April 2007, page numbers (524-533).

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, Vol. 77, Nº 2, page numbers (257-286).

Romero Legarreta, I., Addison, P. S., Reed, M. J., Grubb, N. R., Clegg, G. R., Robertson, C. E. & Watson, J. N. (2005). Continuous wavelet transform modulus maxima analysis of the electrocardiogram: beat-to-beat characterization and beat-to-beat measurement. *Int. J. Wavelets, Multiresolution Inf. Process.* Vol. 3, page numbers (19-42).

Sahambi, J. S., Tandon, S. M. & Bhatt, R. K. P. (1997a). Using wavelet transforms for ECG characterization: an on-line digital signal processing system. *IEEE Eng. Med. Biol.,* Vol. 16, page numbers (77-83).

Sahambi, J. S., Tandon, S. M. & Bhatt, R. K. P. (1997b). Quantitative analysis of errors due to power-line interferences and base-line drift in detection of onsets and offsets in ECG using wavelets. *Med. Biol. Eng. Comput.,* Vol. 35, page numbers (747-751).

Sahambi, J. S., Tandon, S. M. & Bhatt, R. K. P. (1998). Wavelet base ST-segment analysis. *Med. Biol. Eng. Comput.,* Vol. 36, page numbers (568-572).

Sartene, R., et al., (1994). Using wavelet transform to analyse cardiorespiratory and electroencephalographic signals during sleep, In: Proc. IEEE EMBS Workshop on Wavelets in Med. and Biol., page numbers (18a-19a), Baltimore.

Schelkens, P.; Munteanu, A.; Barbarien, J.; Galca, M.; Nieto, X. & Cornelis, J. (2003). Wavelet coding of volumetric medical datasets, *IEEE Trans. Med. Imag.*, Vol. 22, March 2003, page numbers(441-458).

Schiff, S. J., Aldroubi, A., Unser, M., & Sato, S., (1994). Fast wavelet transformation of EEG, In: Electroencephalogr. Clin. Neurophysiol., vol. 91 (6), page numbers (442-455).

Senhadji, L., Carrault, G., Bellanger, J. J., & Passariello, G., (1995). Comparing wavelet transforms for recognizing cardiac patterns, In: IEEE Eng. in Med. and Biol. Mag., vol 14 (2), page numbers (167-173).

Shyu, L-Y., Wu, Y-H. & Hu, W. (2004). Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG. *IEEE Trans. Biomed. Eng. ,* Vol. 51, page numbers (1269-1273).

Simoncelli, E.; Freeman, W.; Adelson, E. & Heeger, D. (1992). Shiftable multiscale transforms, *IEEE Transactions on Information Theory - Special Issue on Wavelet Transforms and Multiresolution Signal* Analysis, Vol. 38, No. 2, March 1992, page numbers (587–607).

Simoncelli, E. & Freeman, W. (1995). The Steerable Pyramid: A Flexible Architecture for Multi- Scale Derivative Computation, Proceedings of IEEE Second International Conference on Image Processing, Washington, DC, October 1995.

Sivannarayana, N. & Reddy, D. C. (1999). Biorthogonal wavelet transforms for ECG parameters estimation. *Med. Eng. Phys.,* Vol. 21, page numbers (167-174)

Strickland, R. N., & Hahn, H. I., (1994). Detection of microcalcifications in mammograms using wavelets, In: *Proc. SPIE Conf. Wavelet Applicat. in Signal and Image Process. II*, vol. 2303, page numbers (430-441), San Diego, CA.

Sung-Nien, Y.; Kuan-Yuei, L. & Huang Y. (2006). Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model, *Computerized Medical Imaging and Graphics,* Vol. 30, No. 3, April 2006, page numbers (163-173)

Tikkanen, P. E. (1999). Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal. *Biol. Cybernetics,* Vol. 80, page numbers (259-267).

Valois, R. De & Valois, K. De (1988). *Spatial Vision,* Oxford Univ. Press, New York.

Vetterli, M. & Kovacevic, J. (1995). *Wavelets and Subband Coding*, Englewood Cliffs, Prentice Hall, NJ.

Wang, K., & Shamma, S. A. (1995). Auditory analysis of spectrotemporal information in acoustic signals. *IEEE Eng. in Med. and Biol. Mag.,* Vol. 14, No. 2, page numbers (186-194)

Watson, A. B. (1987). The cortex transform: rapid computation of simulated neural images. *Computer Vision Graphics Image Process.,* Vol. 39, No. 3, page numbers (311-327).

Watson, J. N., Addison, P. S., Clegg, G. R., Holzer, M., Sterz, F. & Robertson, C. E. (2000). Evaluation of arrhythmic ECG signals using a novel wavelet transform method. *Resuscitation,* Vol. 43, page numbers (121-127).

Watson, J. N., Uchaipichat, N., Addison, P. S., Clegg, G. R., Robertson, C. E., Eftestol, T., & Steen, P.A., (2008). Improved prediction of defibrillation success for out-of-hospital VF cardiac arrest using wavelet transform methods. *Resuscitation,* Vol. 63, page numbers (269-275).

Weaver, J.; Yansun, X.; Healy Jr, D. & Cromwell, L. (1991). Filtering noise from images with wavelet transforms, *Magn. Reson. Med.*, Vol. 21, October 1991, page numbers (288–295)

Xu, J., Durand, L. & Pibarot, P., (2000). Nonlinear transient chirp signal modelling of the aortic and pulmonary components of the second heart sound. *IEEE Transactions on Biomedical Engineering,* Vol. 47, Issue 10, (October 2000) page numbers (1328-1335).

Xu, J., Durand, L. & Pibarot, P., (2001). Extraction of the aortic and pulmonary components of the second heart sound using a nonlinear transient chirp signal model. *IEEE Transactions on Biomedical Engineering,* Vol. 48, Issue 3, (March 2001) page numbers (277-283).

Yang, L.; Guo, B. & Ni, W. (2008). Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform, *Neurocomputing*, Vol. 72, December 2008, page numbers (203-211)

Yang, X., Wang, K., & Shamma, S. A. (1992). Auditory representations of acoustic signals. *IEEE Trans. Informat. Theory,* Vol. 38, (February 1992) page numbers (824-839).

Yi, G., Hnatkova, K., Mahon, N. G., Keeling, P. J., Reardon, M., Camm, A. J. & Malik, M. (2000). Predictive value of wavelet decomposition of the signal averaged electrocardiogram in idiopathic dilated cardiomyopathy. *Eur. Heart J.,* Vol. 21, page numbers (1015-1022).

Yildirim, I. & Ansari, R. (2007). A Robust Method to Estimate Time Split in Second Heart Sound Using Instantaneous Frequency Analysis, *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, pp. 1855-1858, August 2007, Lyon, France.

Zhang, X-S., Zhu, Y-S., Thakor, N. V., Wang, Z-M. & Wang, Z. Z. (1999). Modelling the relationship between concurrent epicardial action potentials and bipolar electrograms. *IEEE Trans. Biomed. Eng.,* Vol. 46, page numbers (365-376).