Proceedings of the 2005 IEEE
Engineering in Medicine and Biology 27th Annual Conference
Shanghai, China, September 1-4, 2005

Optimal discrete wavelet design for cardiac signal processing

J.M.H. Karel[1], R.L.M. Peeters[1], R.L. Westra[1], K.M.S. Moermans[2], S.A.P. Haddad[3] and W.A. Serdijn[3]

***Abstract*— The question of designing the best wavelet for a given signal is discussed from the perspective of orthogonal filter banks. Two performance criteria are proposed to measure the quality of a wavelet, based on the principle of maximization of variance. The method is illustrated and evaluated by means of a worked example from biomedicine in the area of cardiac signal processing. The experimental results show the potential of the approach.**

I. INTRODUCTION

In recent years wavelets have been successfully used in a large number of biomedical applications. The multi-resolution framework makes wavelets into a very powerful compression [5] and filter tool [7], and the time and frequency localization of wavelets makes it into a powerful tool for feature detection [2]. For more advanced applications, however, the success of these techniques depend to a considerable extent on the choice of a good wavelet.

This paper addresses the important question of how to design a good wavelet for an application at hand. Various authors have made suggestions about how to approach this question, see [6], [3], [4]. However, hardly any worked applications are available in the literature. It is the purpose of this paper to describe an approach to this problem based on the theory of orthogonal filter banks and to investigate the performance of this approach for the purpose of cardiac signal processing.

II. ORTHOGONAL WAVELET DESIGN FROM FILTER BANKS

To deal with the question of designing a good wavelet in a systematic way, we employ the theory of wavelets derived from filter banks, see [6]. This provides an elegant and well understood framework for discrete-time (digital) signal processing to which we shall be confined in this paper. There are four important aspects of filter banks to take into account (in decreasing order of importance): (i) perfect reconstruction, (ii) orthogonality of the filter bank and the underlying wavelet based multi-resolution structure, (iii) flatness of the filters and vanishing moments in the wavelets, (iv) smoothness of the wavelets. Once a filter bank has been selected, which corresponds to an underlying *scaling function* $\phi(t)$ and an associated *mother wavelet* $\psi(t)$, the question needs to be addressed how well it is suited to perform its intended signal processing task. This requires a *criterion function*, which allows one to measure the performance and to compare the different filter banks.

[1]Department of Mathemathics, Universiteit Maastricht. P.O. Box 616, 6200 MD Maastricht, The Netherlands. {joel.karel, ralf.peeters, westra}@math.unimaas.nl

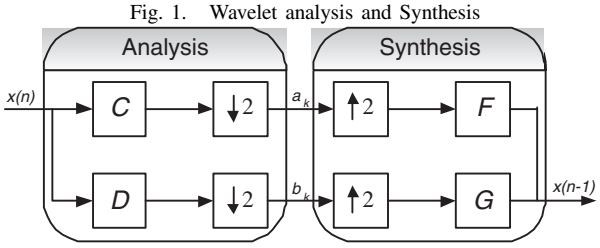
[2]Atlas Copco Airpower n.v., Boomsesteenweg 597, 2610 Wilrijk, Belgium. kurt_moermans@pandora.be. Contributed to this work in his master thesis while studying at Universiteit Maastricht.

[3]Electronics Research Laboratory, Delft University of Technology. Mekelweg 4, 2628 CD Delft, The Netherlands. {s.haddad, w.a.serdijn}@ewi.tudelft.nl

The authors thank Richard Houben of Medtronic BRC for his ideas.

This paper is part of the BioSens project that is being funded by STW (Project no. DTC 6418).

Fig. 1. Wavelet analysis and Synthesis



A. Orthogonal wavelets and filter banks

The idea of filter bank theory is to perform signal processing by means of a bank of (digital) filters in combination with down-sampling. This process is illustrated in Fig. 1. At each stage of the filtering process, the given signal x is passed through a low pass filter with transfer function $C(z)$ and a high pass filter with transfer function $D(z)$ after which down-sampling takes place. These two filters must meet certain requirements to enable perfect reconstruction from the two output signals after down-sampling and to yield an orthogonal underlying wavelet basis. To end up with a corresponding mother wavelet $\psi(t)$ having compact support, $C(z)$ and $D(z)$ must be *finite impulse response* (FIR) filters:

$$C(z) = c_0 + c_1 z^{-1} + \ldots + c_N z^{-N}, \tag{1}$$

$$D(z) = d_0 + d_1 z^{-1} + \ldots + d_N z^{-N}. \tag{2}$$

Here N is an integer determining the order of the filters.

Orthogonality of the underlying wavelet basis involves the following conditions on the filter coefficients: (a) *normalization*: $\sum_{k=0}^{N} c_k^2 = 1$ and $\sum_{k=0}^{N} d_k^2 = 1$; (b) *double shift orthogonality*: $\sum_{k=0}^{N} c_k c_{k-2\ell} = 0$ and $\sum_{k=0}^{N} d_k d_{k-2\ell} = 0$ for all integers $\ell \neq 0$; and (c) double shift orthogonality between the two filters: $\sum_{k=0}^{N} c_k d_{k-2\ell} = 0$ for all integers ℓ. (In the last two conditions negatively indexed coefficients are all zero by convention.)

Double shift orthogonality implies that N is odd, say $N = 2n-1$. The *alternating flip construction* allows one to specify the high pass filter coefficients $d_0, d_1, \ldots, d_{2n-1}$ in terms of the low pass filter coefficients $c_0, c_1, \ldots, c_{2n-1}$ in such a way that condition (c) is automatically fulfilled:

$$d_k = (-1)^k c_{2n-1-k} \qquad (k = 0, 1, \ldots, 2n-1). \tag{3}$$

0-7803-8740-6/05/$20.00 ©2005 IEEE.

It is described in [6, Ch. 4-5] how the remaining orthogonality constraints can be handled by reparameterization of the $2n$ low pass filter coefficients $c_0, c_1, \ldots, c_{2n-1}$ in terms of n new parameters $\theta_1, \theta_2, \ldots, \theta_n$. The theory of polyphase matrices offers a convenient way to achieve this. For $k = 1, \ldots, n$, let $R(\theta_k) = \begin{bmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{bmatrix}$ and let $\Lambda(z) = \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix}$. Then consider the 2×2 matrix product

$$H(z) = \Lambda(-1)R(\theta_n)\Lambda(z)R(\theta_{n-1})\Lambda(z)\cdots R(\theta_2)\Lambda(z)R(\theta_1). \quad (4)$$

Let $H(z)$ be partitioned as $H(z) = \begin{bmatrix} C_0(z) & C_1(z) \\ D_0(z) & D_1(z) \end{bmatrix}$ where

$$\begin{aligned} C_0(z) &= c_0 + c_2 z^{-1} + c_4 z^{-2} + \ldots + c_{2n-2} z^{-(n-1)} \\ C_1(z) &= c_1 + c_3 z^{-1} + c_5 z^{-2} + \ldots + c_{2n-1} z^{-(n-1)} \\ D_0(z) &= d_0 + d_2 z^{-1} + d_4 z^{-2} + \ldots + d_{2n-2} z^{-(n-1)} \\ D_1(z) &= d_1 + d_3 z^{-1} + d_5 z^{-2} + \ldots + d_{2n-1} z^{-(n-1)} \end{aligned}$$

from which $C(z)$ and $D(z)$ are formed as $C(z) = C_0(z^2) + z^{-1}C_1(z^2)$ and $D(z) = D_0(z^2) + z^{-1}D_1(z^2)$.

It then holds that $C(z)$ and $D(z)$ satisfy all the orthogonality constraints mentioned above, as well as the relationship described by the alternating flip construction. Conversely, for all such $C(z)$ and $D(z)$ there exist parameters $\theta_1, \ldots, \theta_n$ which achieve this decomposition of the polyphase matrix $H(z)$.

B. Vanishing moments

Apart from orthogonality, an important desirable property for wavelets is to have vanishing moments. This requires $C(z)$ to have zeros at $z = -1$, thus exhibiting a corresponding degree of flatness at the low and high frequencies (and likewise for the high-pass filter $D(z)$).

To have one vanishing moment amounts to the condition $c_0 - c_1 + c_2 - c_3 + \ldots + c_{2n-2} - c_{2n-1} = 0$ which is equivalent to the commonly imposed condition that the integral of the mother wavelet $\psi(t)$ is equal to zero: $\int_{-\infty}^{\infty} \psi(t)dt = 0$. In terms of $\theta_1, \theta_2, \ldots, \theta_n$ this is well known to translate into the condition $\theta_1 + \theta_2 + \ldots + \theta_n = \pi/4 \bmod 2\pi$. This allows one to build a vanishing moment into the filter bank by simply eliminating one of the free variables.

The conditions corresponding to more vanishing moments are not difficult to obtain in terms of the filter coefficients $c_0, c_1, \ldots, c_{2n-1}$ by requiring also some derivatives of $C(z)$ to have a zero at $z = -1$. They amount to a set of linear conditions in terms of these filter coefficients, but in terms of the parameters $\theta_1, \theta_2, \ldots, \theta_n$ the expressions attain a more difficult form. We state here that the additional condition in case of a second vanishing moment is given by: $\cos(2\theta_n) + \cos(2\theta_{n-1} + 2\theta_n) + \ldots + \cos(2\theta_2 + \ldots + 2\theta_{n-1} + 2\theta_n) + \frac{1}{2} = 0$.

C. Computing $\phi(t)$ and $\psi(t)$

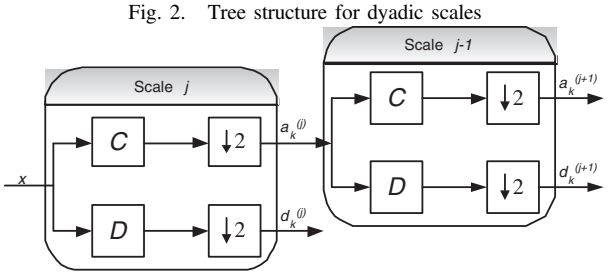
The scaling function $\phi(t)$ is obtained as the solution of the *dilation equation* and $\psi(t)$ then follows from the *wavelet equation*:

$$\phi(t) = \sqrt{2} \sum_{k=0}^{N} c_k \phi(2t - k), \quad (5)$$

$$\psi(t) = \sqrt{2} \sum_{k=0}^{N} (-1)^k c_{N-k} \phi(2t - k). \quad (6)$$

An iteration scheme which allows for the exact computation of $\phi(t)$ at all the dyadic points up to an arbitrary resolution can be found, for instance, in [6]. We have adopted this method in this paper. It is important to note that it may well happen that these functions exhibit a discontinuous and fractal structure.

Fig. 2. Tree structure for dyadic scales



D. Two criteria for wavelet design

For denoising and compression and also for various detection purposes, it is attractive to measure the quality of a filter bank and its underlying wavelet in terms of the way the given signal is represented by the wavelet at various positions and on various scales. For the orthogonal wavelet bases discussed in this paper, this is conveniently described by the wavelet coefficients (and the approximation coefficients) obtained at the various scales by applying the filter bank.

At the finest level, the signal is represented by the time series $x = (x_0, x_1, x_2, \ldots)$ having the total energy (or variance) $E = \sum_k x_k^2$. At the first level, after application of the filters $C(z)$ and $D(z)$ and down-sampling, two sequences of coefficients are obtained: the approximation coefficients $a^{(1)} = (a_0^{(1)}, a_1^{(1)}, a_2^{(1)}, \ldots)$ and the detail coefficients $d^{(1)} = (d_0^{(1)}, d_1^{(1)}, d_2^{(1)}, \ldots)$. Because of orthogonality of the filter bank the total energy is preserved: $E = \sum_k (a_k^{(1)})^2 + \sum_k (d_k^{(1)})^2$. Then at the second level, the filters $C(z)$ and $D(z)$ are applied to the approximation sequence $a^{(1)}$ to produce the next approximation sequence $a^{(2)}$ and the next detail sequence $d^{(2)}$. Again, the total energy remains preserved, and this process continues for as many levels as desired. See Fig. 2.

Since the wavelet coefficients $d_k^{(s)}$ (together with the approximation coefficients at the coarsest level) represent the contribution to the representation of the signal x in terms of the multi-resolution wavelet basis, it is natural to strive for a filter bank which represents the signal x by means of just a few large wavelet coefficients containing most of the energy and many small wavelet coefficients containing only little energy. The L_2-norm is not able to achieve this because of preservation of energy. A guiding principle proposed here is to aim for *maximization of the variance*, either

maximization of the variance of the absolute values of the wavelet coefficients, or maximization of the variance of the squared wavelet coefficients i.e. of the energy distribution over the wavelet contributions at the various scales.

Theorem 2.1: Let $\{w_k | k = 0, 1, \ldots, m\}$ be the sequence of the wavelet coefficients at all the levels and the approximation coefficients at the coarsest level resulting from the processing of a signal $x = (x_0, x_1, x_2, \ldots, x_m)$ by means of an orthogonal filter bank. Then:
(a) Maximization of the variance of the sequence of absolute values $|w_k|$ is equivalent to minimization of the L_1-norm $V_1 = \sum_{k=0}^{m} |w_k|$.
(b) Maximization of the variance of the sequence of energies $|w_k|^2$ is equivalent to maximization of the L_4-norm $V_4 = (\sum_{k=0}^{m} |w_k|^4)^{1/4}$.

Proof. (a) The variance of the sequence of absolute values $|w_k|$ is given by $\frac{\sum_k |w_k|^2}{m+1} - \left(\frac{\sum_k |w_k|}{m+1}\right)^2 = \frac{E}{m+1} - \left(\frac{\sum_k |w_k|}{m+1}\right)^2$ in which E and m are constant. Hence maximization of this quantity is equivalent to minimization of V_1.
(b) The variance of the sequence of energies $|w_k|^2$ is given by $\frac{\sum_k |w_k|^4}{m+1} - \left(\frac{\sum_k |w_k|^2}{m+1}\right)^2 = \frac{\sum_k |w_k|^4}{m+1} - \left(\frac{E}{m+1}\right)^2$ in which E and m are constant. Hence maximization of this quantity is equivalent to maximization of the L_4-norm $(\sum_k |w_k|^4)^{1/4}$. □

The criteria V_1 and V_4 have both been investigated in the worked examples to design wavelets for various signals. When all wavelet coefficients are taken into account and no weighting is applied, both of these criteria tend to produce similar results. However, when only a few scales are taken into account (e.g. by weighting) the results may become quite different: in case of minimization of the L_1-norm, energy tends to be placed in scales not taken into account, whereas in case of maximization of the L_4-norm this does not happen.

III. EXPERIMENTATION

A. Theoretical test

For the first test, as a proof of principle, random testsignals of length 256 were generated. These signals had the property that they are sparse in the wavelet domain. To construct them a wavelet decomposition of a signal of length 256 is taken and a few (1-3) random detail-coefficients are set to 1 and all others to 0. Next the test signal is reconstructed in the time domain with a wavelet filter based on random $\bar{\theta}_2$ and $\bar{\theta}_3$. The corresponding parameter set $\{\bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3\}$ is likely to give the most optimal representation of the test signal with respect to the L_1-norm. If one searches for the optimal $\{\theta_1, \theta_2, \theta_3\}$, it is therefore most likely that this will be $\{\bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3\}$. In experiments it was observed that this is indeed the case, although the optimization may terminate in a local non-global optimum. Next, additive white noise was added to various test signals with a signal-to-noise ratio (snr) equal to 20 dB. Again the original $\{\bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3\}$ were well recovered.

B. Reference signal

As a practical example from biomedicine, a reference signal was created by averaging heartbeats from an episode with ECG signals from Physionets MIT normal sinus rhythm database [1]. This produced a smooth signal that was upsampled to yield the signal displayed in Fig. 3. This signal is used as a typical ECG signal such that the wavelet will be able to capture the common essence of this type of signal. It was used in all the examples below.

Fig. 3. Reference signal

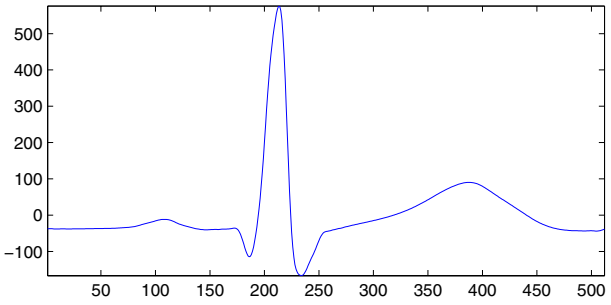
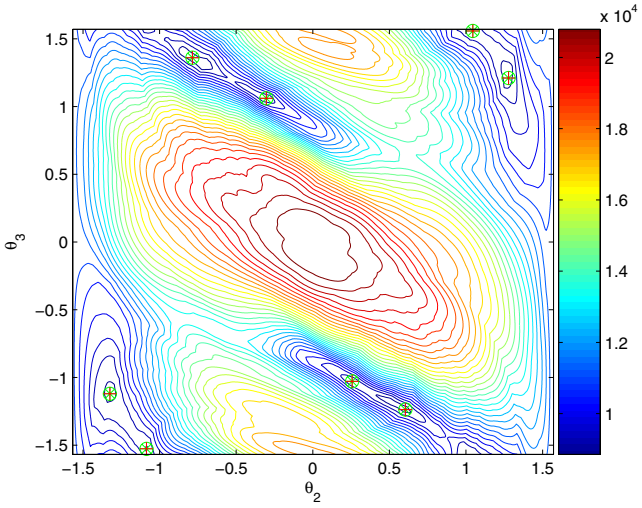


Fig. 4. Local l_1-minima for $n = 3$



C. Local optima

When optimizing a wavelet for the reference signal with respect to a certain norm, the optimization may terminate in a local optimum. In order to gain insight in the existence of these local optima, the situation with $n = 3$, and thus with two degrees of freedom since $\theta_1 + \theta_2 + \theta_3 = \pi/4$, was investigated for L_1-*minimization*. θ_2 and θ_3 are set on a megapixel grid in the range $(-\frac{\pi}{2}, \frac{\pi}{2}]$, as shown in Fig. 4. Some of the local optima have been marked with red/green stars in this figure. When considering some of the filter coefficients corresponding to the local optima, the resemblance to the Daubechies 2 (db2) filter coefficients is evident. Note that to build the db2 wavelet 4 and not 6 filter coefficients are required and that one degree of freedom is effectively unused if $n = 3$. *This observation gives a rationale for the use of the db2 wavelet for cardiac signals.* It was also observed that some of the other filter coefficients resemble the Daubechies 3 wavelet, however with less success.

Fig. 5. Criterium value related to n

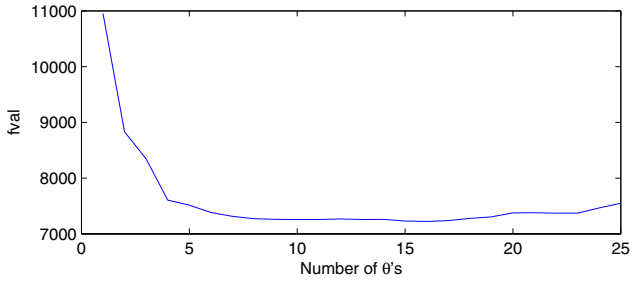
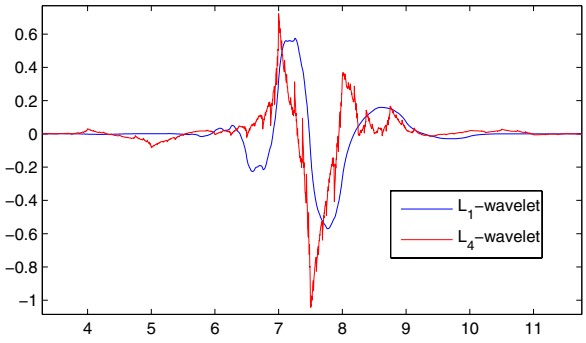


Fig. 6. Wavelets with $n = 8$ optimized for the reference signal



D. Choosing the number n

The filter size of the wavelet filter is determined by the number n. A large n gives more freedom to fit the wavelet to the signal but also increases the complexity. To determine the effect of the choice of n on the criterium value that can be achieved for each $n = 1, \ldots, 25$, 1000 random starting points were generated and the best criterium value in an L_1-sense was stored. See Fig. 5. The results indicate that $n = 8$ is a reasonable choice to work with, which was used in the further experiments below.

E. Practical evaluation

Since the practical application in mind is the use of wavelets in cardiac signal processing, a simple test case was designed to investigate the potential of this approach. As a test set episode 103 of the MIT-BIH arrythmia database [1] was used. This is a 360 Hz annotated, freely available ECG signal. Two wavelets were designed using $n = 8$: one by minimizing the L_1-norm of the wavelet transform of the reference signal, and another one by maximizing its L_4-norm. The two resulting wavelet functions are illustrated in Fig. 6. The wavelet function that was obtained with L_4-maximization has a fractal structure. Despite its fractal nature, the effectiveness of the L_4 optimized wavelet is high, as is illustrated by the wavelet decomposition of the reference signal in Fig. 7. A single strong wavelet coefficient marks the location of largest correlation.

As an experimental setup the wavelet decomposition of the testset with the three wavelets (L_1, L_4 and the popular db2) was calculated, but only a single level (scale) was considered. This level was selected for each wavelet individually to maximize performance. Next a binary vector was constructed of all the wavelet coefficients of which the absolute value exceeded a certain threshold. These locations were then related to locations of the original signal.

Fig. 7. Wavelet decomposition of reference signal with L_4-wavelet

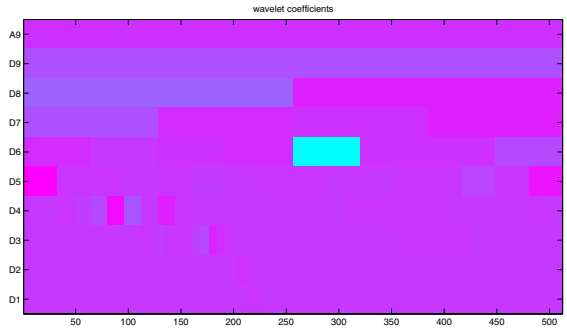


TABLE I
EXPERIMENTAL RESULTS

Wavelet:	L_1-wavelet	L_4-wavelet	db2-wavelet
Level:	3	4	2
Threshold:	35	135	39
True positives:	1682/1688	1688/1688	1687/1688
False positives:	28	3	17

The beat annotations in the testset were used as a reference to locate the QRS-complex. There are 1688 normal QRS-complexes in the testset. If the binary vector corresponding to the wavelet transform has detected a peak within 20 samples (56ms) of the marker, it is assumed that the QRS-complex is detected. If a peak is detected but no marker is within 20 samples, a false detection is assumed. The results of this experiment are illustrated in Table I. This table shows that the performance of the db2 wavelet is quite good. The L_4-optimized wavelet however shows superior performance. Furthermore the L_4-wavelet is more robust with respect to the choice of threshold value, which may be a large advantage in practical applications.

REFERENCES

[1] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, and H.E. Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, June 2000.

[2] Cuiwei Li, Chongxun Zheng, and Changfeng Tai. Detection of ECG characteristic points using wavelet transforms. *IEEE Trans. Biomed. Eng.*, 42(1):21–28, January 1995.

[3] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[4] N. Neretti and N. Intrator. An adaptive approach to wavelet filter design. In *Proc. IEEE international workshop on neural networks for signal processing*, September 2002.

[5] Ivo Provazník and Jiří Kozumplík. Wavelet transform in electrocardiography - data compression. *International Journal of Medical Informatics*, 45(1-2):111–128, June 1997.

[6] Gilbert Strang and Truong Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.

[7] M.P. Wachowiak, G.S. Rash, P.M. Quesada, and A.H. Desoky. Wavelet-based noise removal for biomechanical signals: a comparative study. *IEEE Trans. Biomed. Eng.*, 47(3):360–368, March 2000.