Wavelet analysis of the a-wave of the human electroretinographic signal

R. Barraco, L. Bellomonte, M. Brai

Dipartimento di Fisica e Tecnologie Relative e Sez. INFM Università di Palermo

lbellomo@difter.unipa.it

Abstract: We here analyse the onset of the a-wave of the human electroretinographic signal by means of the wavelet transform which permits the time-frequency representation of the signal. The purpose is to reveal the transient frequency content of the a-wave in relation to the luminance of the light stimulus. The a-wave is processed by the Haar and the Mexican hat mother wavelets. They provide complementary information, the Haar is sensitive to the changes in signal slope, whereas Mexican hat is directly affected by the amplitude of the a-wave. Both evidence the asymmetry of the a-wave and its dependence on the luminance. The results are presented as isometric lines in a time-frequency diagram. They show that the frequency content is negligible at the beginning and at the end of the wave, but it is relevant (and constituted by low frequencies) in the central zone of the wave.

Introduction

The electroretinogram (ERG, Figure 1) is the recording of the retinal response to a light pulse. It consists of a temporal sequence of components (a-, b-, c-, d-wave, late potentials) arising from various retinal layers.

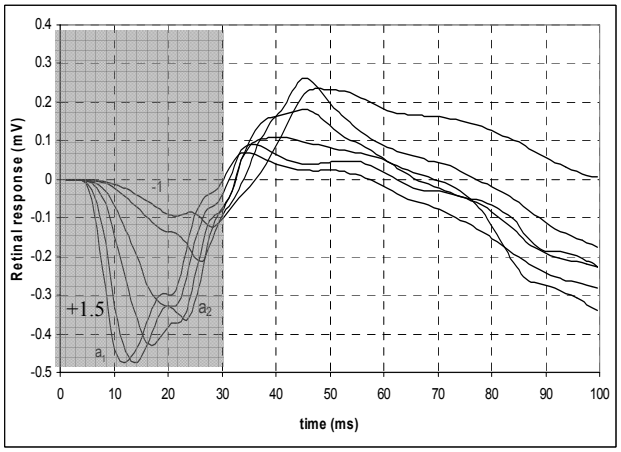

Figure 1: Human ERG recorded in the range of $\log(I/I_0)$ from +1.5 to -1 with step -0.5 ulog (see text) and truncated at 100 ms. The shaded area indicates the time interval of interest.

Understanding the specific features (onset, time delay, amplitude, line shape and so on) of ERG components and their relationships is the principal aim of the research in ocular electrophysiology. We are interested in the time-frequency content of the first 31.25 ms of the a-wave, it appears in the ERG as a small negative potential characterised by two dips, labelled as a_1 and a_2 in Figure 1. They are attributed to the contribution of the photoreceptoral activities of the cones and rods, respectively [1-3]. The relative intensities of their responses and times of occurrence depend on the luminance as evident from Figure 1.

Since the retinal photoreceptors have responses and activation times not uniform in time, the response frequency components have time dependent features. The study of this aspect, useful in understanding the phenomena underlying the early steps of phototransduction, can be partially accomplished by the Fourier analysis. A more sophisticated approach involves the wavelet analysis (WA), able to describe the time and frequency characteristics of a non stationary signal [4-6], it is applied here to the photoreceptoral response.

To investigate the a-wave, we have used two mother wavelets (Haar and Mexican hat, HW and MHW, respectively). The results are reported in Figures 4-9 as isometric lines that reveal the changes in the wavelet transform related to the variation of the value and frequency content of the temporal signal.

Main features of the temporal signal

The incomplete knowledge of both the temporal cut-off of the a-wave and the onset of the following b-wave complicates the study of the a-wave ERG component. In conditions of high luminance (curve labelled 1.5), the dip a_1 occurs at about 13.1 ms, whereas a_2 occurs at about 21 ms and is less pronounced. At intermediate luminance the two dips tend to coalesce into one dip only. At low luminance, the second dip tends to predominate with respect to the first one. The characteristic times of both increase with decreasing the luminance.

Main features of the Wavelet Analysis

The application of the mathematical transforms to a generic signal is very useful for obtaining information not directly deducible from its original temporal diagram. In fact, the description of a signal in the time domain is not always sufficient for its accurate analysis, since significant information about its frequency content is usually hidden. The Fourier transform is appropriate

ISSN: 1727-1983

for the analysis of stationary signals, but it is not adequate for that of biological/medical data. In fact, in these cases, it is important to know not only the values of the component frequencies but also their times of occurrence. The WA provides a time-frequency representation of the signal through the use of functions (wavelets) time-and/or frequency-localised. It uses windows of different size and time location for computing the various frequency components. It is thus possible to identify the constituent frequencies and to determine their temporal features. The WA is, hence, a powerful tool to describe the dynamics of complex non-linear processes characterised by interactions in the space-time framework.

In order to be classified as a wavelet, a function $\psi(t)$ must satisfy certain criteria.

1-It must have a finite energy.

2-The admissibility condition must hold. It implies that the negative and positive areas under the curve must cancel out.

3-The average value of the transform $\psi(\omega)$ must be equal to zero.

These properties imply that the wavelet must be limited in time.

Wavelets satisfying condition 2 can also assimilated to band pass filters. They allow the transmission of those signal components which fall within a finite range of frequencies accepted by the filter.

The wavelet transform returns a data vector of the same length of the input data. The analysing wavelet is first multiplied by the a-wave signal, then the coefficients are calculated from the evaluation of the area under the resulting curve. The area volumes can be plotted in the time-scale domain providing the three-dimensional representation of the signal. The wavelet transform is defined according to

$$WT_x^{\psi}(\tau, s) = \sum_{t=0}^{N-1} x(t)\psi * \left(\frac{t-\tau}{s} \right) \delta t \qquad (1)$$

In (1), x(t) is the signal, $\psi^*[(\tau - \tau)/s]$ is the complex conjugate of the mother wavelet, N is the number of samples constituting the signal, τ indicates the temporal delay of the wavelet and s the width of the window, δt is the sampling time (time resolution of the signal). The translation parameter τ and the scale s are related to the time and to the inverse of the frequency band, respectively. High scales (global views) correspond to coarse frequencies resolution, whereas low scales (detailed views) to fine frequency resolution. If a signal has a repeated pattern, and if the pattern of the wavelet is similar, the WT coefficients assume extremal values when both overlap. Therefore, WT may be considered as a measure of similarity between the two signals. It provides sufficient information about the analysis and synthesis of the original signal.

In computational practice, the wavelet transform is computed for a limited range of scales, conditioned by the sampling interval and the number of samples. At the beginning, the most compressed wavelet is placed at $\tau = 0$. It is displaced (by varying τ) it is dilated (by increasing s), the procedure is repeated until the largest dilatation ($s = N\,\delta t/2$) is reached. To each value of s and τ there corresponds a point in the translation-scale plane, in a three-dimensional plot, the ordinate is the value of the transform.

In an ideal situation, every wavelet should cover a single point in time and a single point in frequency. This aim cannot be achieved since if one wants to know a signal at a certain instant of time, nothing can be said about its local frequency behaviour and vice versa, this is analogous to the Heisenberg uncertainty relation in quantum mechanics. The WA gives, hence, good time and poor frequency resolution at high frequencies, good frequency and poor time resolution at low frequencies. To small scales there correspond small time uncertainness appropriate to detect high frequency components characterised by a rapid time variation. Slowly varying signals, on the other hand have low frequency components and large windows are appropriate.

There exist a variety of wavelets. The particular choice depends on the signal to be processed and on the requested information. Selection of a wavelet shape will show a part of the raw signal with specificity. In the actual case, the a-wave has a near gaussian shape and its average temporal width is about 18.5 ms, the half width of the Fourier transform is about 35Hz. Consequently, its frequency content falls in a large band, and a wavelet transform suitable to separate the frequency components, such as the Morlet transform is not appropriate. We have, hence, chosen two real wavelets (Figure 2) that supply a good time frequency resolution: the Haar (step function) and the Mexican Hat (second derivate of a gaussian).

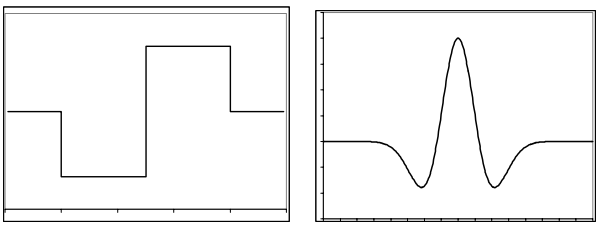

Figure 2: Left: Haar function; right: Mexican hat

They differentiate in the information can be gathered. The first one has the characteristic of being conceptually simple, fast, symmetric and, orthogonal, but it has the disadvantage of being discontinuous and therefore not differentiable. It is well localized in time but the wings of its Fourier spectrum determine a bad localization in the frequency domain. The transform output is affected by the variation of the slope of the temporal signal: if the slope is constant in the time interval covered by the wavelet, the output is null; if it varies, the output of the transform is proportional to this variation (in value and sign).

The MHW is, on the other hand, a continuous wavelet, not orthogonal, defined by (second derivative of a Gaussian):

ISSN: 1727-1983

$$MHW(t) = \frac{1}{\sqrt{2\pi}}(1-t)^2 e^{-\frac{t^2}{2}} \qquad (2)$$

The output of the MHW transform is directly conditioned by the value of the temporal signal, it is hence suitable to isolate local minima/maxima of the signal at the selected time. A constant d-c shift of the signal changes the output that depends on the amount of the area delimited by the signal, regardless of its shape. A change in sign of the signal causes a change in the output. With respect to the Haar transform, time localization of the MHW is poorer, but frequency localization is improved.

Materials and methods

ERG acquisition

The study was carried out on 10 subjects (20 eyes) not affected by ocular diseases, with normal ERGs, visual acuity of 20/20 (range ±2) and negligible differences between the two eyes. ERGs were recorded following our routine methods [7] by means of Henke's corneal electrodes. The indifferent electrodes were frontal, with the ground electrode on the forehead. Electrode impedance was kept under 5 kΩ. Stimuli were stroboscopic white flashes of 50 μs presented in a Ganzfield integrating sphere of 40 cm diameter. The standard luminance, denoted I_0, was 1.7 cd*sec/m^2, it was varied in the range +1.5 ulog, –3.5 ulog by neutral Kodak Wratten filters in steps of -0.5 ulog. In order to semplify the labels in the figures, we have indicated the variable luminance I in terms of log (I/I_0). The repetition interval of the stimulus was 15 s and its duration was 50 μs. All subjects were submitted to maximal papillary dilatation (≥ 7mm) with N-ethyl-(α-picolyl) tropicamide 1%, and the cornea was anaesthetised using oxibuprocaine hydroclorate 4%. They were then dark-adapted for 30 minutes, in accordance with the standards for clinical electroretinography [8]. The lower luminance limit (denoted –3.5) was reached, with steps of 0.5 ulog, using neutral Kodak Wratten filters.

Original data was sampled with a frequency of 1,024 Hz and stored as ASCII files in the mass memory for subsequent retrieval and analysis. Each clinical analysis consisted of 11 traces (each containing 512 values covering the time interval of 500 ms.). Each trace was obtained by averaging at least 3 responses.

Computational approach

The sampling interval δt was set equal to 0.97563 ms. Each record contains 32 values corresponding to 31.25 ms. In the WA the smallest scale was set $s_0 = 2\delta t$ and the largest one $s_{Max} = 32\delta t$. The corresponding frequency range is 32–512 Hz. The temporal translation was in the range 0–31.25 ms with step of 0.97563. Since we were dealing with time delimited signals, errors may occur in the computation of the wavelet transforms for large windows located at the temporal extremes of the signal. The periodic conditions were adopted in order to avoid anomalies in the results. The wavelet amplitude intensity is usually represented by a surface in a three-dimensional space. The figures, presented here, are bi-dimensional plots showing the projections of these surfaces, on the plane (s, τ) as iso-lines. The red/blue areas correspond to positive/negative values of the transform, respectively, the areas in the green-yellow range represent intermediate values. The upper horizontal axis reports the scale in ms from 1 up to 16 (linear with step 3), the lower one the corresponding frequencies in Hz from 512 up to the 32.8, logarithmic. The vertical axis corresponds to the translations of the window related to the parameter τ, by the coefficient 0.97563. The characteristic features of the transform are represented by the contours shapes of the isometric lines. The contours of the iso-lines have been smoothed to simplify the reading of the results.

Results and Discussion

We here limit the discussion to three representative a-wave corresponding the values of log(I/I_0) equal to +1.5, +0.5, -0.5. The relative traces are reported in Figure 3.

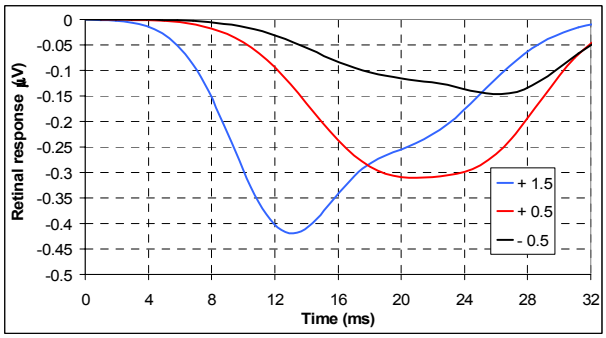

Figure 3: Temporal behaviour of the a-wave recorded at log(I/I_0) equal to +1.5, +0.5, -0.5 in the range of 32 ms. The curves have been smoothed to simplify the reading.

It is evident the dip dependence on the luminance and its shift toward greater times as the luminance is reduced.

It is convenient to separate the discussion of the results of the two transforms although at least one aspect is commons to both: asymmetries in the signal produce deformed contours lines, in different manner in the two transforms.

Haar Transform

Figures 4-6 report the trend of the iso-lines of the Haar transform relative to the values of log(I/I_0) equal to +1.5, +0.5, -0.5, respectively. Two regions characterised by nearly elliptic concentric iso-lines are evident: the blue one occurs at smaller times and corresponds to negative variations of the slope, the red one indicates positive slope and occurs at greater times. These zones

ISSN: 1727-1983

are separated by a green area that indicates values of constant slope, these are also found at the beginning and at the end of the analysis time where the signal is slowly variable. The regions of more intense colour indicate the times of the greatest variation of the temporal signal slope.

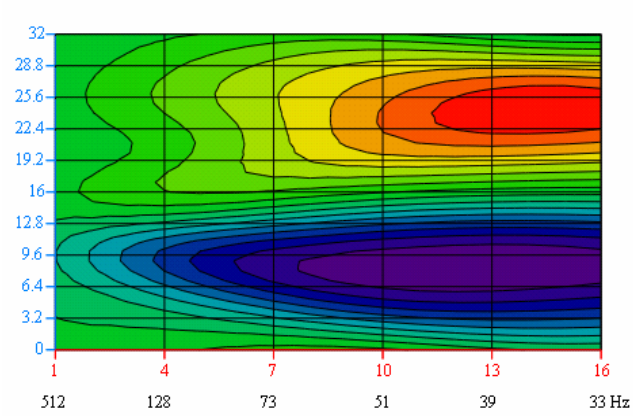

Figure 4: Haar transform applied to the a-wave recorded at the largest luminance ($\log(I/I_0) = +1.5$). The abscissa corresponds to the scale (red) in ms and the correspondent frequency bandpass in Hz (black), the ordinate indicates the locations of the window, to have the delay time (in ms), one must multiply by 0.97563.

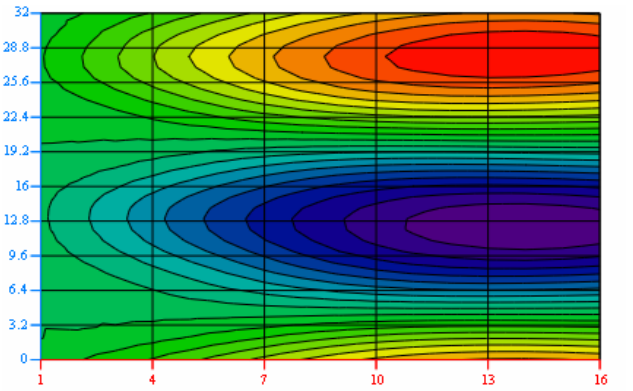

Figure 5: As Figure 4 with $\log(I/I_0) = +0.5$

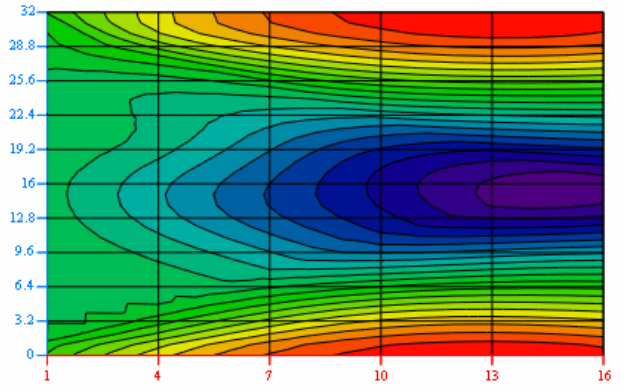

Figure 6: As Figure 4 with $\log(I/I_0) = -0.5$

In Figure 4, the left apex of the blue area, at small values of the window occurs at 9.6 ms, it corresponds to the time of occurrence of the largest temporal variation of the signal and moves slowly towards smaller values of time as the width of the window increases. This is related to the fact that the slope is not constant and large windows comprehend large portions of asymmetric signal. However, the contour is somewhat regular as consequence of the fact that in the time interval covered by a generic window, the signal is monotone. At greater times, the green area takes place indicating that the temporal signal is slowly variable. The signal then tends to increase and the area switches toward red colours, at small values of the window a bifurcation is evident since an inflexion region is present between the two dips. The two apexes occur at 16 ms and 25.6 ms, respectively. As the window becomes greater and greater, the two dips may fall inside the same window, the resolution becomes poorer and poorer, the bifurcation tends to disappear and the two apexes coalesce into one at the intermediate time of about 24 ms. At this value of the scale, the time uncertainness is so high that the dips are smeared out.

Figures 5 and 6 display a similar trend with some relevant differences: i) the iso-lines move toward higher times as a consequence of the luminance reductions, ii) their shapes change since the temporal signal deforms, in particular for $\log(I/I_0) = +0.5$, the two dips coalesce, it follows that the bifurcation in Figure 5 disappears, whereas for $\log(I/I_0) = -0.5$, (Figure 6) the two apexes reappear with some differences.

This fact is in line with the experimental evidence that, in the temporal signal, the relative intensity of a dip with respect to the other one depends on the value of the luminance.

Mexican hat Transform

Figures 7-9 report the trend of the isolines of the Mexican hat transform relative to the same values of $\log(I/I_0)$.

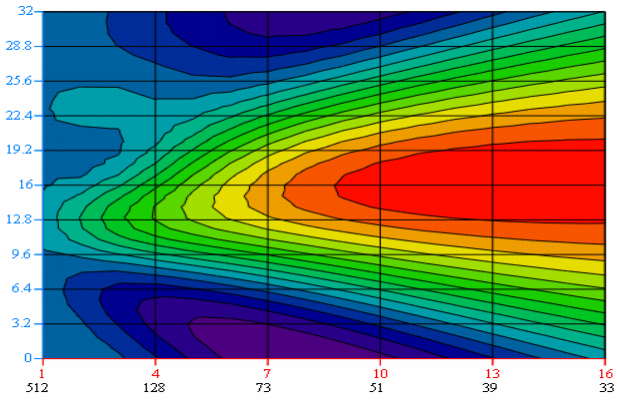

Figure 7: As Figure 4. Transform calculated using the MHW.

ISSN: 1727-1983

The central zone in Figure 7 is constituted by iso-lines that cover areas in the green – red range. The green region corresponds to small values of the scale, the amount of signal captured by the window is small, but the resolution is good and the two dips are resolved. The first is located at values of τ of 12.8 ms, the second at 22.4 ms, these correspond to the times of occurrence of the two dips in the temporal signal. As the window broadens the amount of signal falling inside it increases, smaller frequencies enter into the region delimited by the iso-lines determining the colour switch toward the red. The resolution becomes poorer and the dips are no longer resolved and a single apex (at $\tau \approx 16$ ms) appears as s is greater than 8.5ms.

The deep blue areas, corresponding to values of the transform close to zero, are ascribed to the cancellation among the positive contributions (product of the signal and of the MHW) and the negative ones in the calculation of the transform.

Figures 8 and 9 yield a similar behaviour with the difference, already mentioned, that the temporal signal deforms and moves toward reater times. In particular,

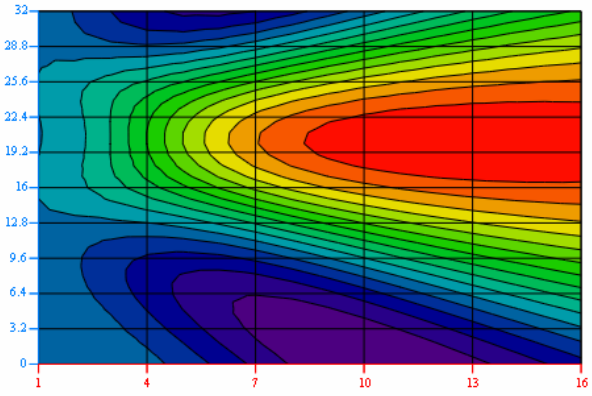

Figure 8: As Figure 5. Transform calculated using the MHW.

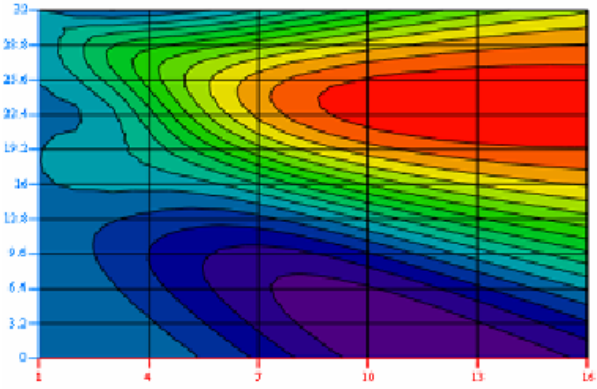

Figure 9: As Figure 6. Transform calculated using the MHW.

the tendency of two dips to coalesce, appears in Figure 8 as an apex flattening in the central zone. The translation toward higher values of τ must be attributed to the asymmetric shape of the temporal signal.

Confidence interval

The confidence interval is defined as the probability that the true wavelet power, $\psi^2(\tau,s)$, at a certain time and scale, lies within a certain interval about the estimated wavelet power. The confidence interval for $\psi^2(\tau,s)$, is defined according to:

$$\frac{1}{\chi^2\left(\frac{p}{2}\right)}\left|\Psi\ (\tau,s)\right|^2 \le \psi^2(\tau,s) \le \frac{1}{\chi^2\left(1-\frac{p}{2}\right)}\left|\Psi\ (\tau,s)\right|^2 \quad (3)$$

where p is the chosen significance (p=0.05 for the 95% confidence interval) and $\chi^2(p/2)$ represents the value of χ^2 at p/2. Using (3) one can then determine the confidence intervals for the peaks in a wavelet power spectrum to compare them with the mean background or against other peaks.

Conclusions

The present results demonstrate the usefulness of the wavelet analysis to individuate frequency transients in a non-stationary signal such as the human a-wave. This applies to Figures 4-6 as well as to Figures 7-9.

In fact, from the analysis of the first set of figures regarding the Haar wavelet, we note that the value of this transform is conditioned by the slope of the signal. In the temporal intervals in which the slope is appreciably variable, the evaluation of the transform returns high values either positive or negative. They yield the occurrence of a large amount of frequency components in which the low frequencies predominate.

A complementary information can be deduced from the analysis of Figures 7-9 which evidence the content and the time of activation of the various frequency components constituting the temporal signal. At the beginning and at end of the wave, this content is small, vice-versa in the time interval covering the central part, a large amount of oscillation modes give a relevant contribution. This is large for low-frequency modes and decreases as long as greater frequencies are considered. The amount of involved frequencies is weakly dependent on the luminance, whereas the times of activation are strongly conditioned by the luminance. It appears that, the low frequency content is less affected by the luminance than the higher one. This fact is related to the fast shape variation of the signal at times close to those of occurrence of the two dips and to their dependence on the luminance.

It is interesting to relate the above considerations with the analysis of the Fourier spectra of the signals (Figure 3) reported in Figure 10. These agree with the results of Figures 4-6 and 7-9 for what concerns the

ISSN: 1727-1983

frequency content, but they are unable to reveal the temporal distribution of the various frequencies. The predominance of the low frequencies can also be explained in the context that the shape of the temporal signal approximates that of a gaussian, whose frequency spectrum is characterised by a large amount of low frequency components.

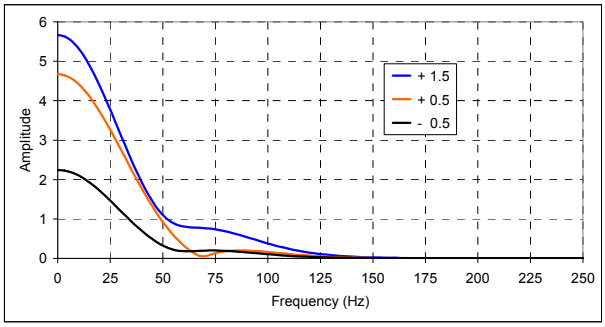

Figure 10: Spectral distribution of the signals plotted in Figure 3. The ordinate reports the modulus of the Fourier transform. The units are arbitrary, but the factor of proportionally is the same.

For completeness, it is useful to report the frequency spectra of the two wavelets used here. See Figures 11 and 12.

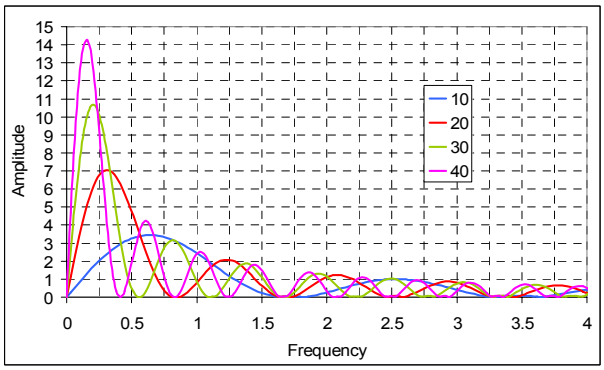

Figure 11: Spectral distribution of four Haar functions characterised by different widths in arbitrary units.

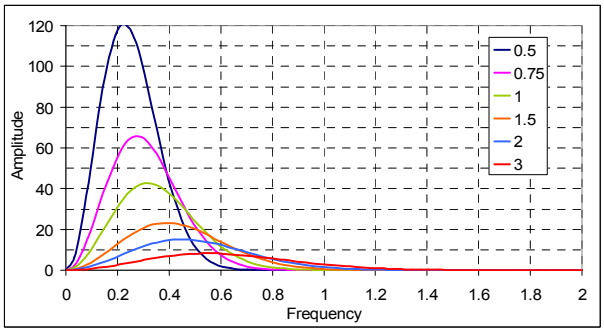

Figure 12: Spectral distribution of six Mexican hat functions characterised by different widths in arbitrary units.

These spectral behaviours are quite different, even for different window widths. They have in common the null value of the transform at the zero frequency, this is a consequence of the fact that the area enclosed by the two wavelets is null.

The spectral distribution of Haar function is characterised by various damped oscillations related to the temporal discontinuity of the wavelet, it follows that it covers a wide frequency interval. In the Mexican Hat spectrum, on the other hand, only one peak is present. In both cases the spectral width broadens as the width of the wavelet decreases.

From the previous considerations, it turns out that the present approach provides a suitable compromise between time domain and frequency domain localisation of a signal.

The application of wavelet transforms in electrophysiology is a relatively new field of research, therefore, some aspects such as the choice of an appropriate mother wavelet, scale parameters and delay times, needs a further investigation to enhance the use of this signal processing method. In particular, it is auspicated that controversial aspects concerning the typology of the processes underlying the early retinal response may be clarified.

References

[1] Burns M. E., Lamb T. D. Visual Transduction by Rod and Cone Photoreceptors in Chalupa L. M. and Werner J. S. (Eds): 'The Visual Neurosciences' 2003: 215-33.

[2] Hood D., Birch D. G. A quantitative measure of electrical activity of the human rod photoreceptors using electroretinography. Visual Neuroscience 1990: 5, 379-387.

[3] Hood D, Birch D G. The relationship between models of receptors activity and the a-wave of the human ERG. Clin. Vision Sci. 1990: Vol. 5 N° 3 293-297.

[4] N M Astaf'eva. Wavelet analysis: basic theory and some applications. Physics-Uspekhi Fizicheskikih Nauk 1996: 39, 1085-1108.

[5] Rioul O., Vetterli M. Wavelet and signal processing. IEEE SP Magazine 1991 14-38.

[6] Torrence C., Compo G. P. A pratical guide to wavelet analysis. Bullettini of the American Meteorological Society. 1998; 71, 1; 61-78.

[7] Anastasi M., Brai M., Lauricella M., Geracitano R. Methodological aspects of the applications of the Naka-Rushton equation to clinical electroretinogram. Opthalmic. Res. 1993: 2, 145-156.

[8] International Standardisation Committee. Marmor et al. Standard for clinical electro-retinography. Arch ophthalmol 1989; 107, 816-819.

ISSN: 1727-1983