Technical University of Cluj-Napoca
Faculty of Electronics, Telecommunications and Information Technology

APPLICATION OF B-SPLINE WAVELET ANALYSIS IN BIOMETRICS

PhD Thesis Summary

Graduand: Eng. Maria Simina MOLDOVAN (EMERICH)
Scientific advisor: Prof. Dr. Eng. Corneliu RUSU

Cluj-Napoca
2011

1. Introduction

The theory behind wavelets has been developed during the last thirty years independently by mathematicians, scientists and engineers. Researchers are faced with an ever increasing variety of wavelet bases to choose from. The first purpose of the thesis was to research the properties of B-spline wavelets (fractional, generalized, Battle-Lemarie, biorthogonal) and to study their significant impact on the early development of the wavelet transform. Further, the thesis proposes the improvement of several wavelet analysis techniques by using B-spline functions family.

From the very beginning of wavelets, it has been recognized the strong connection between wavelets and differential operators. We propose to investigate how it can be further improved on the wavelet's behavior as differentiator using the B-spline wavelet transform. The spectrogram and scalogram allow the examination of the energy distribution in the time-frequency respectively scale-frequency plane. Energetic analysis is connected to the chosen type of wavelet family. We suggest a comparison among B-spline wavelets energetic results and the ones acquired by other well known wavelets. In addition, a multiresolution framework for analysis of glottal closure instants is proposed.

A privileged area of applications where wavelet methods have been found to be relevant is pattern recognition. Biometric recognition is adverted to the automatic identification of a person based on some specific vectors, derived from the biological characteristics. In the present thesis three biometric systems are implemented (based on dynamic signature, voice and iris).

2. The improvement of the wavelet analysis techniques by using B-spline functions family

Wavelet analysis is a relative new method, mathematically based on the work of Joseph Fourier. Many signals contain transitory characteristics which are often the most important part and Fourier analysis could not detect them. In 1946 D. Gabor adapted the Fourier transform to analyse only a small section of the signal at a time. Gabor's adaptation, called the Short Time Fourier Transform (STFT), maps a signal into a two-dimensional function of time and frequency. STFT has limited precision, determined by the size of the analysis time window. Wavelet analysis represents the next logical step: a windowing technique with variable-sized regions. It allows the use of long time intervals for precise low-frequency information and shorter regions for high-frequency information. While the choice of the best wavelet is application depend, it seems to be very useful to isolate a number of features thet are of great interest to the users. Unlike most other wavelet bases, splines have explicit formulae both in time and frequency, which explain their significant impact on the theory of the WT. As more and more wavelet solution are proposed, the selection of a particular wavelet should be motivated by the problem itself. Comparative studies are needed more then ever.

Polynomial splines with uniform knots were first introduced by Schoenberg in 1946. Fractional B-spline functions were proposed in 2000 by T. Blu and M. Unser. The primary motivation for considering fractional B-splines instead of conventional ones was that the enlarged family happens to be closed under fractional differentiation. The uniform first order B-spline function is defined as:

$$\beta_+^0(x) = \Delta_+^1 x_+^0 \quad \xleftrightarrow{F} \quad \hat{\beta}_+^0(w) = \frac{1-e^{-jw}}{jw} \begin{matrix} -\textit{finite} & \textit{difference} \\ -\textit{derivative} & \textit{operator} \end{matrix} \tag{1}$$

where $x_+^\alpha = \max\{0, x\}^\alpha$ is the one-sided power function, $\Delta_+ = \delta(x) - \delta(x-1) \overset{F}{\leftrightarrow} 1 - e^{-jw}$ is the finite difference operator and $\partial \overset{F}{\leftrightarrow} jw$ is the derivative operator. The *k*-th order function is defined by convolving the *k-1* order function with the first one. Analogically with the classical B-splines, the new family is constructed using linear combinations of the integer shifts of the one-sided power functions:

$$\beta_+^\alpha(x) = \frac{\Delta_\alpha^{\alpha+1} x_+^\alpha}{\Gamma(\alpha+1)} \quad \xleftrightarrow{F} \quad \hat{\beta}_+^\alpha(w) = \left(\frac{1-e^{-jw}}{jw}\right)^{\alpha+1} \tag{2}$$

where α is the fractional degree of the B-splines function and $\Gamma(\alpha)$ is the gamma function.

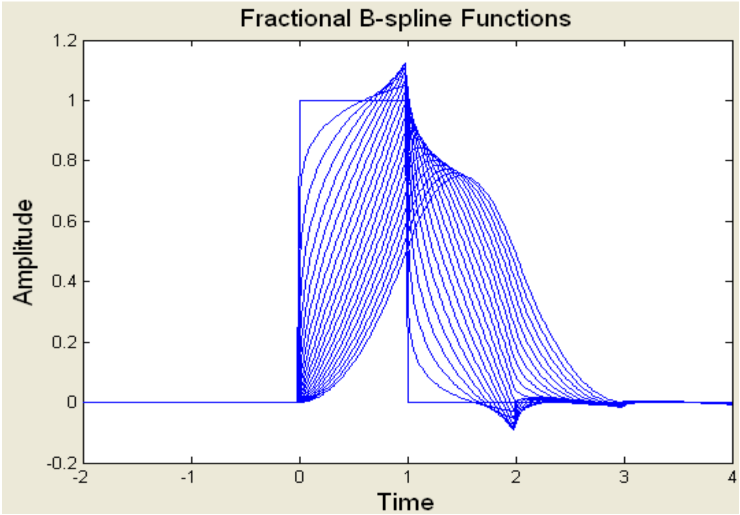


Figure 1. The Fractional B-spline Functions of 0; 0.1; 0.2; … ; 1.9; 2 orders

As we mentioned before, one of the primary reasons for the success of B-splines in applications is their **derivative-like behaviour**. This property generalizes nicely to the fractional case:

$$\partial^{\gamma}\beta_{+}^{\alpha}(x) \xleftrightarrow{F} (jw)^{\gamma}\left(\frac{1-e^{-jw}}{jw}\right)^{\alpha+1} = \left(1-e^{-jw}\right)^{\gamma}\left(\frac{1-e^{-jw}}{jw}\right)^{\alpha+1-\gamma} \tag{5}$$

$$\partial^{\gamma}\beta_{+}^{\alpha}(x) = \Delta_{+}^{\gamma}\beta_{+}^{\alpha-\gamma}(x) \tag{6}$$

The fractional derivative of order γ of an α fractional B-spline is another $\alpha-\gamma$ fractional B-spline. The key operator in this case is the causal fractional finite difference one: $\Delta_{+}^{\alpha+1} \xleftrightarrow{F} \left(1-e^{-jw}\right)^{\alpha+1}$. In our study, we also appeal to generalized B-spline functions, proposed by Van De Ville. His starting point was the fractional B-spline family but he added the shift parameter (τ). The fractional derivative of order (α',τ') of a (α,τ) generalized B-spline is another $(\alpha-\alpha',\tau-\tau')$ generalized fractional B-spline.

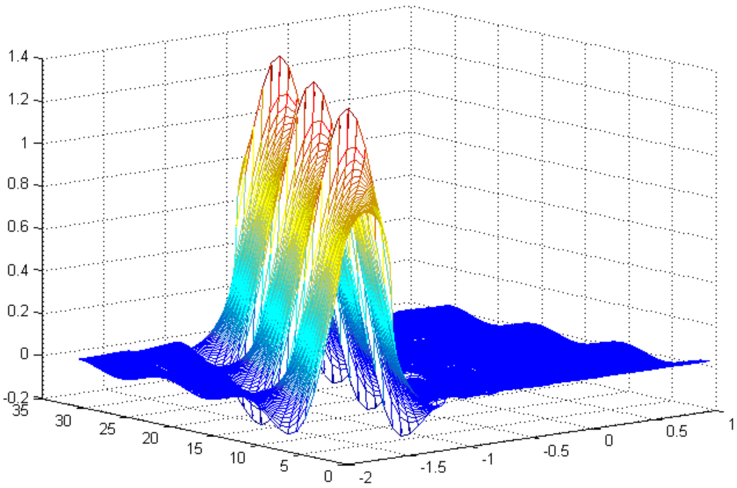


Fig.2. Generalized fractional B-spline ($\alpha = 0.5,\ \tau = -0.5; -0.4; -0.3; \ldots;\ 2.4; 2.5$)

Stationary Wavelet Transform (SWT) is designed to overcome the lack of translation-invariance of the discrete wavelet transform (DWT). The combination of SWT with optimal wavelet functions can be regarded as a smoothing and a differentiation process.

Figure 3 presents the detail coefficients obtained from the first level for different wavelet techniques of a piecewise polynomial signal (a). The first two decompositions (b) and (c) are obtained by using the *'db1'* wavelet function, for SWT respectively DWT method. The third decomposition results from WT based on the generalized fractional B-spline function for $\alpha = 0.4$ and $\tau = 0.2$. Signal singularities are compactly characterized by SWT and by WT used with the fractional B-spline function. This feature opens up new ways to analyze a signal using B-spline wavelet bases.

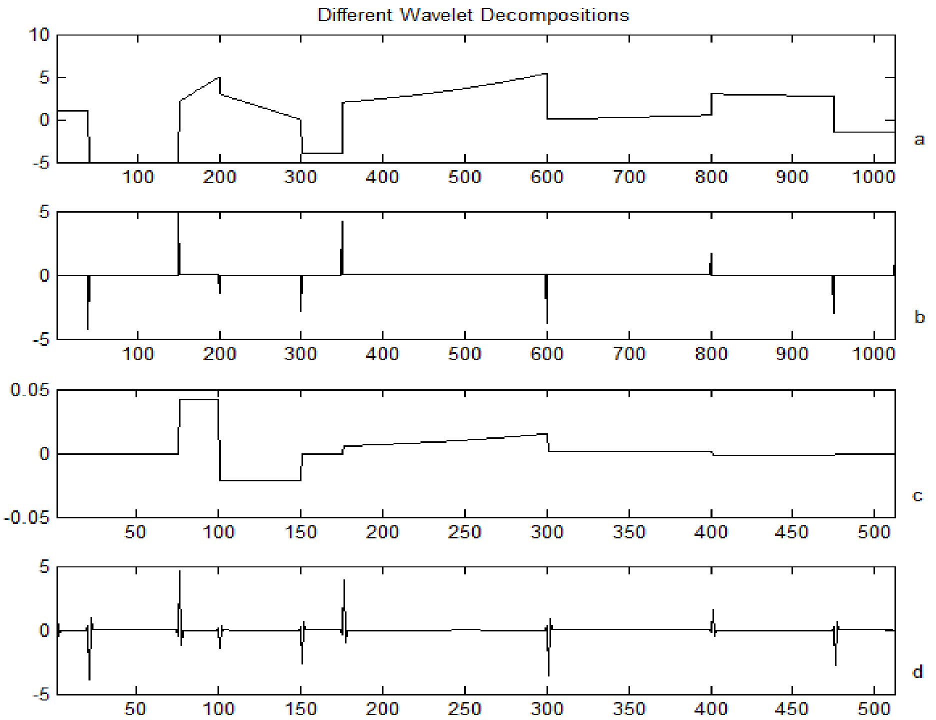
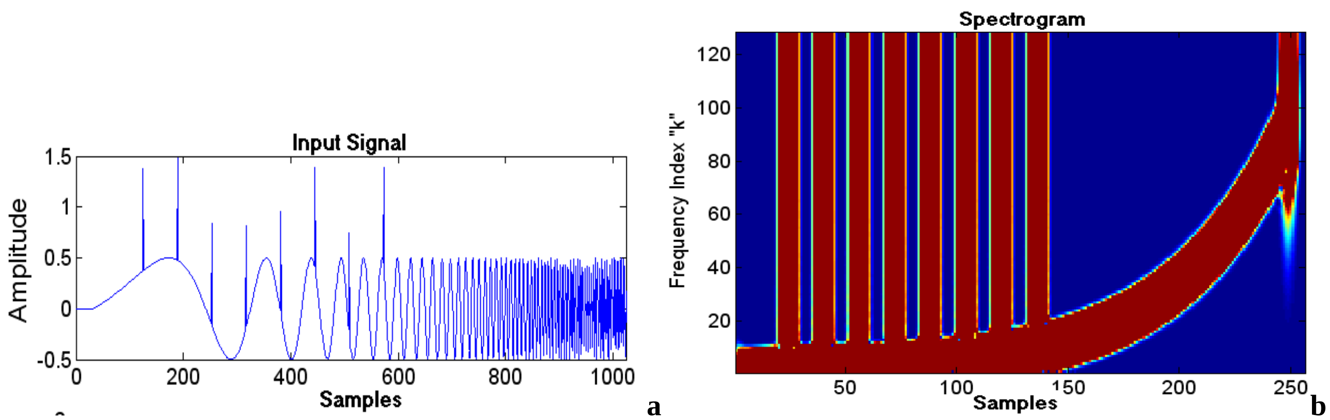


Figure 3. a. Input signal; b. SWT – first level detail coefficients, *db1*; c. DWT – first level detail coefficients, *db1*; d. WT – first level detail coefficients, generalized B-spline *(0.4,0.2);*

Energetic analysis of signals -connected to the chosen type of wavelet function

The Fourier spectrogram is defined as the square modulus of the STFT. If the length of the window is small, the spectrogram will be well localized in time, but it will have poor frequency resolution and vice versa. WT has become a valuable analysis tool due to its ability to elucidate simultaneously both the spectral and the temporal information within the signal. Wavelet spectrogram, called scalogram, communicates the time frequency localization property of the wavelet transform.

Figure 4.b. presents the spectrogram resulted from considered entry signal: an exponential chirp with 8 added transients. Figures 4.c and 4.d illustrate the resulted scalograms using dB3 and the fractional B-spline function of 0.7 order (blue - low energy; red - high energy). The energy distribution region of an impulse located at $t=t_i$ for the spectrogram is limited on the time scale by the time frame and it expands uniformly over all frequencies. On the contrary, for the scalogram the energy is only represented for certain frequencies and it is concentrated in the vicinity of t_i on a time scale. Therefore, the frequency content of the impulses and also their time localizations is more accurate in the case of scalogram. It clearly offer a better visualization for both the transients and the chirp. We have constructed the test waveform so that the frequency increases exponentially starting with the 50th sample. This can be seen only from the second spectrogram. In the beginning, and particularly in the final bi-dimensional representation, on both scalogram and spectrogram some errors could arise due to the “zero padding” phenomenon. In the case of the fractional B-spline wavelets, these errors are reduced.



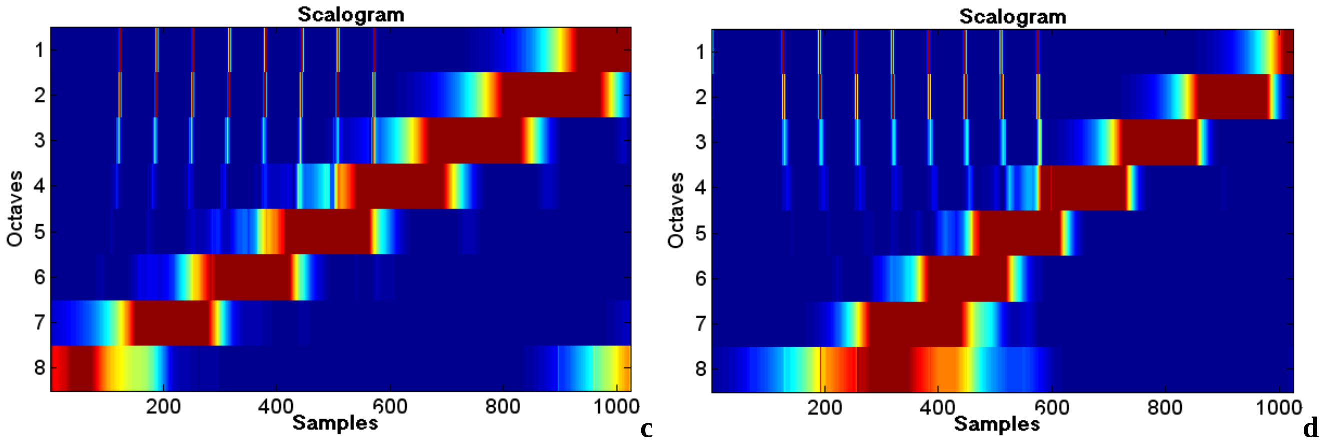


Figure a. Input Signal; b. Spectrogram; c. Scalogram - dB3; d. Scalogram - fractional B-spline

The Estimation of Glottal Closure Instants in Voiced Speech

The glottal source waveform is an important characteristic used in voice analysis, speaker emotional state identification, speech synthesis etc. The amplitude changes in the speech signal can be related with the glottal wave's phases. The thesis proposes to express these amplitude changes in successive scales and search for local maxima corresponding to the closed glottis instants (GCI). In this purpose a multiresolution analysis is performed. Beyond the analysis of different signals, using the Daubechies, Shannon and fractional B-spline wavelets, further observations could be made: In the case of using fractional B-splines there are several energy levels of different intensities which are distributed on several octaves. Many details revealed to us, and even more information was made available besides that provided by the other wavelet functions.

We present some results obtained by applying the fractional B-spline wavelet filter bank and by applying Frobenius norm. Figure 5.b. represents the 4th octave of the wavelet decomposition computed by using the fractional B-spline wavelet of 0.4 order. Figure 5.c. represents the Frobenius measure for GCI detection. The input signal is also mixed with a white noise - SNR=10dB (Figure 6.a) and is analyzed by using the same methods. Determination of the instants of glottal closure from speech wave using wavelet transform is equivalent to finding a particular local modulus maxima pattern across several scales in the time-scale plane. When detecting the GCI, the octave band decomposition showed superior performance in comparison with the covariance methods. It can be seen that the Frobenius measure fails completely in noisy conditions and the multiresolution detector remains stable.

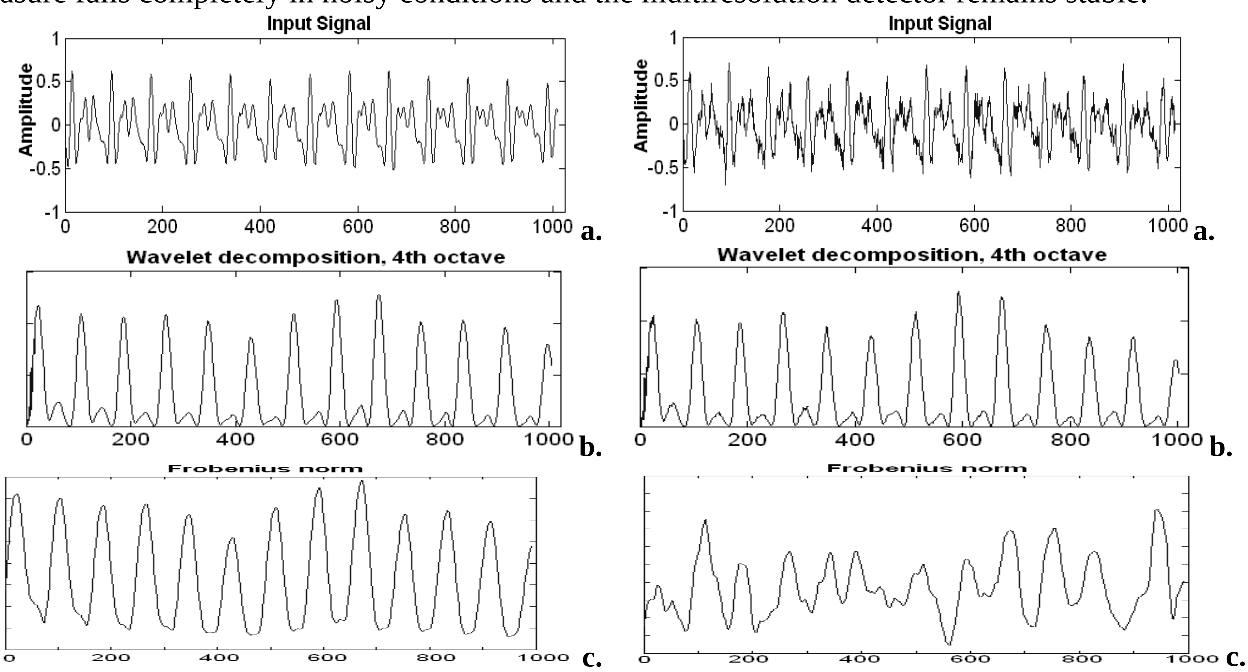


Figure 5. a. Voice signal; b. WT, 4th octave, B-spline 0.4; c. The Frobenius norm.

Figure 6. a. Voice signal + white noise; b. WT, 4th octave, B-spline 0.4; c. The Frobenius norm.

3. Biometric Systems

The thesis proposes different biometric systems based on dynamic signature, voice and iris. For the experiments some classifiers (kNN, Naive Bayes and SVM) provided by the WEKA (Waikato Environment for Knowledge Analysis) environment are employed. In the feature extraction process different wavelet techniques are adapted and combined with the modified TESPAR DZ coding method.

The idea behind the TESPAR (Time Encoded Signal Processing and Recognition) method is the employment of an approximation model based on the zeros theory. The waveform is divided in periods determined by successive passes through zero of the signal, thus maintaining the time information combined with a simple approximation of the waveform between two successive passes through zero. The thesis employs a version of this method based on TESPAR DZ matrices. In this case three descriptors are used to describe every epoch: D (*duration*), S (*shape*) and A (*amplitude*). Applying TESPAR DZ procedure, pairs of epochs are compared; a symbol is produced indicating the differences between the individual D, S and A features of the two epochs being compared.

Biometric System Based on Dynamic Signature and Speech

Signing is part of everyday life and is perceived as a non-invasive process by the majority of the users. On-line or dynamic signatures are acquired by a graphic tablet. The voice is usually employed by persons to recognize each other during the dialog carried from long distance. Speech as biometric verifier has some important advantages: the low price for the sensors (microphone), the noninvasive way of acquisition and the ability to provide real time processing. Speech and dynamic signature signals present more similarities which allow processing them by similar techniques

A public available handwritten on-line signature database (Task2-SVC2004) was employed in this research. The corpus consists of 40 sets with 40 signatures for each user (20 genuine signatures and 20 skilled forgeries). We also built our own biometric bimodal database (BimDB10) involving speech and signature traits from 10 individuals, women and men (7 Romanian and 3 French). For each user 100 signatures/utterances have been registered, during 5 different sessions.

Feature Set Extraction for Dynamic Signature –The Stationary Wavelet Transform (SWT) and WT based on fractional B-splines is performed on the selected time functions of the signature: *X-coordinate*, *Y-coordinate* and *Pressure*. We extract the approximation coefficients of the first level of decomposition (cA1) in order to de-noise the signals and the details coefficients of the first three levels (cD1, cD2, cD3). A zero crossing in the detail coefficients usually corresponds to a peak or valley in the input signal. In the case of approximation signals, each epoch is characterized by its duration D, amplitude A and shape S, while in the case of detail signals each epoch is characterized by its duration D and amplitude A. Comparisons between consecutive epochs were made in the present study. For each individual epoch pair, a two-stage vector, instead of a three-stage one, as in the classical method is generated for each descriptor in signature analysis. Consequently, the dimensionality for each aproximation coefficients set is reduced to 8 and for each details set it is reduced to 4 due to the modified TESPAR DZ technique. In addition, 4 wavelet energy coefficients were used (one coresponding to the approximation and 3 for details). Thus, a vector of a prefixed dimension was obtained for every signature – Sign_{64}. Issued features are independent of the size and location of the signatures. This independence was achieved by extracting them with translation and rotation invariant techniques based on wavelet and TESPAR combination.

Feature Set Extraction for Speech –By applying the TESPAR DZ coding procedure 27 coefficients are obtained from each password signal. Other, the speech signals are decomposed by using the Perceptual Wavelet Packet Transform. Such a PWPT is designed to match the psychoacoustic model. The sampling rate is 16 kHz, yielding a speech bandwidth of 8 kHz. Within this bandwidth, there are approximately 24 critical bands as shown in Figure 7. For each sub band~~,~~ the mean energy is calculated. Additional features resulted from the time analyses are also used: relative mean square energy, number of maxima in the mean energy envelope, mean pitch frequency and normalized zero crossings rate. A fixed length vector is obtained for each utterance – Voice_{55}.

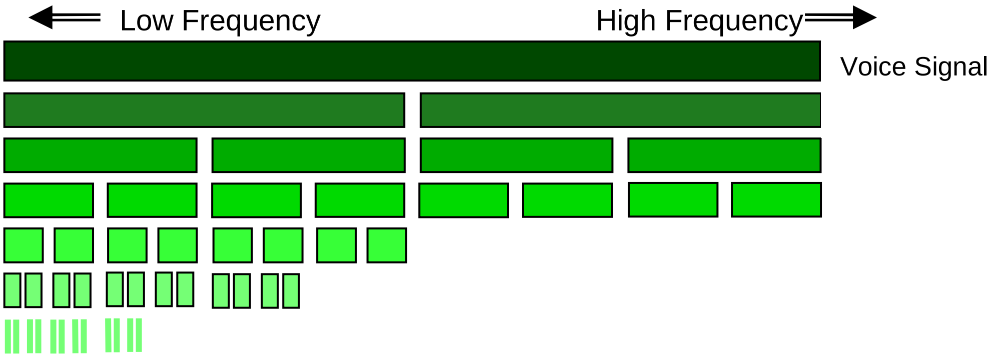


Figure 7. The proposed wavelet packet decomposition

Experiment Set 1- Signature Identification Task

We carried out identification experiments, using 20 signatures per person (20x40 instances for SVC2004 database respectively 20x10 for BimDB10). The SVM classification performances are tested for all available kernels (linear, polynomial, RBF and sigmoid). Researchers are faced with an ever increasing variety of wavelets to choose from and the choice of the best wavelet is application-dependent. We select several well known wavelet functions such as dB1, dB3, lem1, lem2, bior1.3, bior2.4, rbio1.3, rbio2.4 and coif1. Additionally, the behaviour of fractional B-spline wavelets ($\alpha = 0$, 0.2, 0.4, 0.7, 1) and generalized B-spline wavelets ($\alpha = 0$, 0.2, 0.4, 0.7, 1 and $\tau = 0.2$) was investigated. The purpose was to obtain a robust model and to use it further for the verification task. Lem1, together with the RBF kernel (C=100 and $\gamma = 0.01$), is indicated by our results (93.26% classification rate) and will be use for further experiments. Good performances are also obtained for Db1, fractional B-splines ($\alpha \to 0$) followed by bior1.3. The common aspect of these function is that they act like first order derivative. In fact, applying SWT on the X-coordinate and Y-coordinate, respectively, we get the velocity (Vx, Vy) as first level details, and the acceleration (Ax, Ay) as second level details. WT based on these functions can be regarded as a smoothing and a differentiation process, yielding a robust and performant system. Compared to the conventional methods, there are some advantages in using SWT for derivative calculation, e.g. simplicity in algorithm and improvement in SNR.

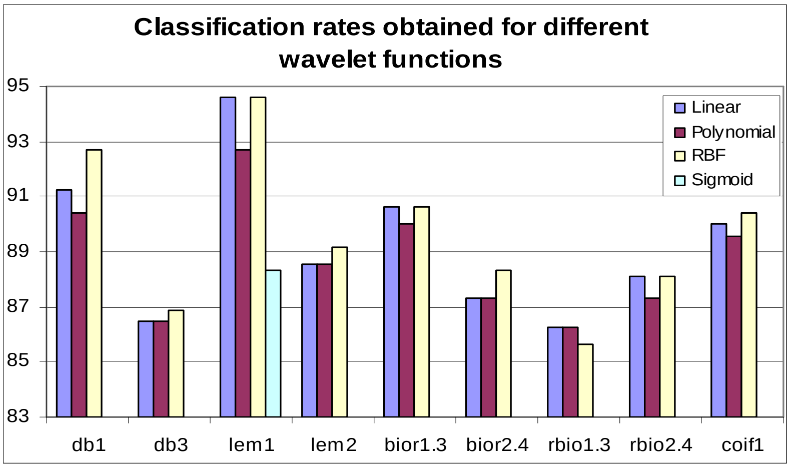


Figure 8. Identification experiments, SVC2004 database

Experiment Set 2 - Signature Verification Task

The use of skilled forgeries is very important to behavioural biometrics such as signature. Verification is the decision about whether the signature is genuine or forgery. Each set contains 20 genuine signatures/ signer and 20 skilled forgeries collected from other 5 signers. The technique was based SVM classifier, RBF kernel, lem1 wavelet function. The average accuracy was 93.12%.

Experiment Set 3- - Speaker Identification Experiments

Our experiments reveal that the accuracy rates in the case of speech are not very sensitive to the wavelet functions type. Hence, lem1 will be used in order to preserve the connection with the signature based system. In this case a classification rate of 94.84% was obtained, for the SVM classifier, RBF kernel (C=100, $\gamma = 0.01$), BimDB10_Speech subcorpus, 10x100 signatures.

Experiment Set 4 - Feature Vector Selection and Fusion Experiments

The proposed bimodal system employes the evidence presented by two biometric sources (on-line signature and speech). 5% of the data was used for training (50 signatures/utterances) and the rest of 95% for testing (950 signatures/utterances). The feature selection was made on the test set by using the ChiSquaredAttributeEval method. From the unimodal systems based on BimDB10_Sign and BimDB10_Voice subcorpus, we retained different individual feature vectors and we combined them by using two fusion methods: at the feature level and at the score level.

Feature Vector \ SVM Kernel	Linear	Polynomial	RBF	Sigmoid
Sign64	93,26%	93,15%	93,26%	93,15%
Sign35	94,52%	94,21%	94,84%	94,21%
Sign20	92,31%	89,15%	91,78%	89,15%
Voice55	94,21%	92,31%	94,84%	92,31%
Voice20	94,31%	87,57%	90,84%	87,57%
Voice15	94,31%	87,05%	86,63%	87,05%
Sign35+Voice20	99,89%	99,78%	99,89%	99,89%
Sign20+Voice15	**99,36%**	**99,68%**	**99,36%**	**99,68%**

Table 1. Identification experiments, BimDB10 database

Voice and signature data present complementary information. When these two modalities are fused, the performance and the robustness of the biometric system are improved. For a small size vector ($Sign_{20}$+$Voice_{15}$) a very good accuracy was obtained, above 99%. In the case of Signature based system the subset extracted from x and y coordinate is more relevant than the one extracted from pressure. Since pressure is invisible, it is difficult to forge. In the case of voice based system, TESPAR DZ coding procedure seems to offer very good packing properties.

The performances of the proposed system allow us to mention that the feature extraction and selection steps satisfy the next crucial requirements: intraclass variance is small (meaning that features derived from different samples of the same class are close) and interclass separation is large (features derived from samples of different classes differ significantly).

For verification task, the results obtained from the reduced feature vectors (Sign20, Voice15 and Sign20+Voice15) are reported. For each user, from 50 instances used for learning, 5 were genuine and the rest (9x5) were considered forgeries. The testing dataset consisted of 45 original instances and 9x45 forgeries. The performances of these experiments are expressed in the terms of FRR (false rejection rate), FAR (false acceptance rate) and ROC (Receiver Operating Characteristic). Best results were obtained for the bimodal system, feature vector fusion: both FAR and FRR were under 4%.

Biometric System	FAR	FRR	Verification Rates
Voice	11,05%	12,10%	88,42%
Signature	9,78%	8,10%	91,05%
Bimodal System (feature fusion)	3,26%	3,89%	**96,42%**
Bimodal System (score fusion)	3,36%	4,42%	96,10%

Table 2. Verification experiments, SVM classifier

Biometric system based on iris

Iris recognition is relatively young method, existing in patent only since1994. The employed database is a public one and contains 3 x 128 iris images provided from 64 persons (left and right eye). The first stage of iris recognition is to isolate the actual iris region in a digital eye image. The iris region is approximated by two circles, one for the iris/sclera boundary and another, for the iris/pupil boundary. The circular Hough transform was employed to deduce the radius and centre coordinates of the pupil and iris regions. The detected iris region is unwrapped by remapping each point to a pair of polar coordinates using the cartesian to polar transform suggested by Daugman. In this way, we obtain a rectangular representation for the iris, useful to make things easier for the further processing.

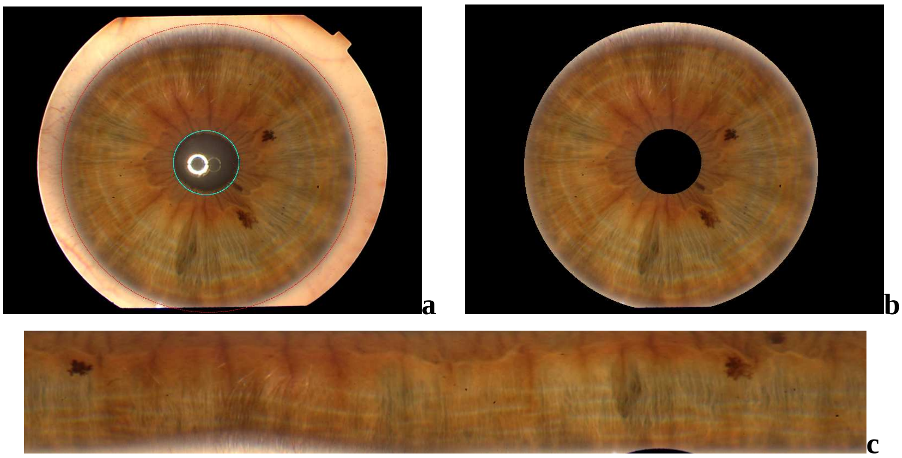


Figure 9. a. Test image having the iris and the pupil detected; b. Cropped iris ROI; c. Unwrapped iris

For iris feature set extraction DWT 2d was applied to the unwrapped iris image, by using one level decomposition. The matrix corresponding to the approximations (cA) and details (cH, cV, cD) were represented as vectors (cA1d, cH1d, cV1d and cD1d). The resulted waveforms could be divided in epochs whose length, amplitude and shape change over time. Consequently the TESPAR DZ coding method could be applied. From the TESPAR DZ histogram of different iris images provided from different user could be seen that 9 from 27 symbols were never detected (the amplitudes of two consecutive epochs are never the same). This situation was eliminated and thus the dimension of the feature vector is 4x18 for each iris. Further, the results obtained for identification experiments in terms of accuracy are presented. For training were used 2 images and for testing 4 images/ user. The SVM classifier, polynomial kernel (C=200 d=7), together with B-spline functions seem to be the best suited.

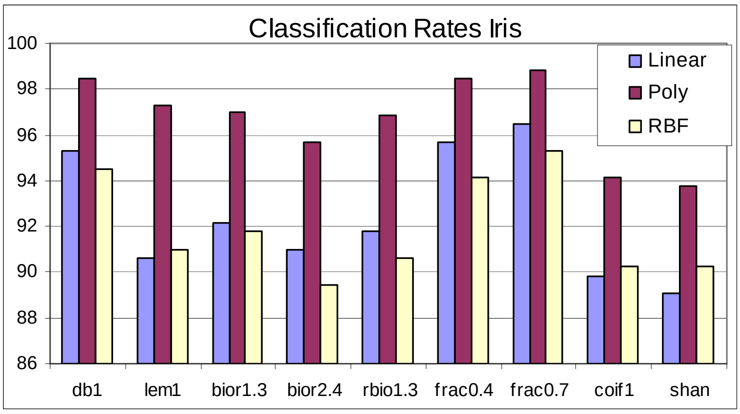


Figure 10. Iris Identification Rates (SVM)

For verification two classes were considered: class A with iris images derived from user A and class nonA with iris images derived from other users. For training were used 2 images class A and 10 images class nonA and for testing 4 images class A respectively 20 images class nonA. Verification experiments were made for 15 users from 64, SVM classifier, polynomial kernel, fractional B-spline wavelet ($\alpha = 0.7$). The accuracy was above 93% for all users. We mention that the idea of using TESPAR DZ method on images is original and were not mention by other researchers.

Author's Contributions

The major contributions brought by this PhD thesis are:

- The study of the B-splines significant impact on the early development of the wavelet transform
- The implementation of the fractional B-spline wavelet filterbanks
- Searching the connection between energetic analysis and the chosen type of wavelet family
- The Estimation of Glottal Closure Instants in voiced speech
- The implementation of three biometric systems based on signature, speech and iris
- The construction of the BimBD10 database.