Multiple Audio Watermarking With Spectrum Pyramid

***Roumen Kountchev*[1], *Mariofanna Milanova*[2], *Vladimir Todorov*[3] and *Roumiana Kountcheva*[3]**

[1]*Faculty of Communication Techniques and Technologies, Technical University of Sofia, Bulgaria*
[2] *Computer Science Department, UALR, Little Rock, USA*
[3] *T&K Engineering, Sofia, Bulgaria*

Summary*:*
The paper presents a new approach for intellectual property protection of audio products when they are distributed via computer networks or radio broadcasting. The presented approach is based on a method for digital watermarking, performed in the spectrum domain, using a Complex Hadamard Transform (CHT) with a specific matrix, in which only ¼ of the elements are complex numbers. In correspondence with this method the watermark data modifies the phases of selected spectrum coefficients of the audio signal. The values of these coefficients are used to build the Spectral Pyramid (SP), which is the synonymous presentation of the processed audio signal. The pyramid permits multi-layer watermarking, because the watermarks, embedded in its layers do not interchange. The paper presents the algorithms for the audio watermark embedding, extraction and detection. The method is evaluated on the basis of some theoretical analysis comparisons and on experimental results. The new approach presents applications in e-commerce, distance learning, tracing and proving of the illegal distribution of audio products.

Key words
Digital watermarking, Spectrum pyramid, Complex Hadamard Transform

1. Introduction

In correspondence with the up-to-date methods for audio watermarking [1-6], the watermarks are inserted in the time- or frequency domain of the audio signal, using a variety of masking effects, which concern the sound perception in accordance with the Human Auditory System (HAS) [8]. To make the distortions resulting from the watermark insertion in the time domain smaller [7,10,11], the watermark is presented as a pseudo-noise binary sequence, added to the corresponding discrete values of the processed audio fragments. When the watermark is embedded in the audio signal spectrum [3,9], the amplitudes and the phases of selected complex low-frequency coefficients are modified (by) using one of the widely known discrete linear transforms: Fourier, Fourier-Mellin, Radon, Hadamard [12], etc. Another approach was developed for discrete cosine and wavelet transforms [4,6], for which the corresponding spectrum coefficients are real numbers. In this case, either a modulation of selected cosine coefficients from the middle-frequency band of the audio signal is used, or a pseudo-random sequence is inserted in some components of its wavelet decomposition. The basic qualities of the spectrum approach for watermark insertion [1,2,9] are its practical inaudibility, the high resistance against pirates' and fraud attacks, the different compressions, time scaling, amplitude corrections, the linear and nonlinear filtration, etc. In the paper is offered a high-efficient method for audio watermarking, based on the modification of the spectrum coefficients' phases in every level of their Spectrum Pyramid, obtained with Complex Hadamard Transform (CHT) [13]. In section 1 are presented the basic methods for audio watermarking; section 2 comprises the mathematical description of the method and the corresponding algorithms for audio watermark embedding, extraction and detection; section 3 contains the evaluation of the watermark efficiency and some results of the method modeling; and section 4 presents the conclusion.

2. Mathematical Description
2.1 Audio Watermark Embedding Based on SP with CHT

The Spectrum Pyramid (SP) as used for digital audio signal decomposition can be explained by first dividing the discrete audio signal in fragments, which consist of same number of samples. The processing starts with amplitude normalization of the fragment data, which is necessary only if the highest amplitude of a fragment sample is smaller than the maximum. Then the audio fragment is processed with the selected linear orthogonal transform in order to build its Spectrum Pyramid. For this, each fragment is processed with CHT, using only a part of the low-frequency transform coefficients. Their values comprise the first (lowest) pyramid level. The fragment is restored from these values with inverse CHT, and after that the restored fragment is subtracted sample by sample from the original one. This "difference" fragment, which is of the same size as the original, is divided in 2 parts and each is processed with the same transform again; the values of the selected coefficients

Manuscript received February 5, 2007
Manuscript revised February 25, 2007

with higher frequencies build the second pyramid level. The calculation of the next pyramid levels' coefficients continues in a similar way. The retained transform coefficients for every pyramid level can be different. In particular, the CHT decomposition of an audio fragment with length $N=2^n$ and samples x(k), used for the calculation of the two-level SP with layers l = 0 and 1, is defined with the following relations:

$$x(k)=\tilde{x}_0(k)+\tilde{e}_0^s(k)+e_1(k)\text{ , where } k=0,1,..,N-1. \quad (1)$$

The component $\tilde{x}_0(k)$ is described for the ***level l=0*** with the approximating model of x(k), defined with:

$$\tilde{x}_0(k)=CHT^{-1}\{F_0[y_0(u)]\},$$
$$y_0(u) = CHT[x(k)] \text{ for } k, u = 0,1,..,N-1. \quad (2)$$

The operators CHT[•] and $CHT[\bullet]^{-1}$ represent the direct and the inverse CHT of the signal samples x(k):

$$y(u)=\sum_{k=0}^{N-1}x(k)t(u,k)=y_r(u)+jy_{Im}(u) \text{ for } u = 0,1,.., N-1,$$
$$x(k)=\frac{1}{N}\sum_{u=0}^{N-1}y(u)t^*(u,k), \quad (3)$$

where $y_r(u)$ and $y_{Im}(u)$ are correspondingly the real and the imaginary part of the complex transform coefficient y(u), represented as a two-dimensional vector in the complex plane:

$$y(u)=M(u)\exp[j\varphi(u)]=M(u)\cos\varphi(u)+jM(u)\sin\varphi(u).$$

Here M(u) and φ(u) represent the polar coordinates of the vector y(u), corresponding with the amplitude- and the phase-frequency spectrums of the audio fragment x(k) (k=0,1,..N-1). In (3) the terms t(u,k) and $t^*(u,k)$ are the elements of the matrices which represent correspondingly the direct / inverse CHT:

$$t(u,k)=j^{-uk}h(u,k) \;;\; t^*(u,k)=j^{uk}h(u,k), \quad (4)$$

$$h(u,k)=\begin{cases}1 & \text{for } n=2;\\ \prod_{r=3}^{n}(-1)^{\lfloor u/2^{r-1}\rfloor\lfloor k/2^{r-1}\rfloor} & \text{for } n=3,4,...\end{cases} \quad (5)$$

Here $\lfloor * \rfloor$ is an operator, representing the integer part of the result, obtained after the division. The analysis of the values of j^{uk} in the terms (4) for $u,k=0,1,..,2^n-1$ (taking into account the relations $j^{4k}=1$, $j^{4k+1}=j$, $j^{4k+2}=-1$, $j^{4k+3}=-j$) shows that 1/4 of the elements t(u,k) are $\pm j$, and the remaining - correspondingly ± 1. In result, the number of the complex multiplications, necessary for the CHT implementation, is lower than the one, used for the known unified CHT [12], which has equal number of elements $\pm j$ and ± 1.

For the pyramid ***level l=1*** at the beginning is defined the difference $e_0(k)=x(k)-\tilde{x}_0(k)$, which is used for the calculation of its approximation model:

$$\tilde{e}_0^s(k)=CHT^{-1}\{F_1[y_1^s(u)]\},\; y_1^s(u)=CHT[e_0(k)] \text{ for } s=1,2. \quad (6)$$

The components $\tilde{e}_0^s(k)$ and $y_1^s(u)$ are defined from the difference $e_0(k)$, for k=0,1,..,(N/2)-1 and s=1. For s=2 these components are calculated for $e_0(k)$, when k=(N/2),(N/2)+1,..,N-1. The operators $F_0[.]$ and $F_1[.]$ represent the filtration of the fragment spectrum in SP levels l=0,1, which is performed using the selected part of the spectrum coefficients only. The retained coefficients should be even number, complex-conjugated couples. In the inverse CHT the "truncated" (removed) coefficients are substituted with zeros. The last, residual component in the decomposition, presented with (1) is defined as:

$$e_1(k)=e_0(k)-\tilde{e}_0^s(k)\text{ , for } s=1,2. \quad (7)$$

Assuming, that from the fragment spectrum in levels l=0,1 are retained the complex-conjugated coefficients with frequencies $u_1=4m+1$ and $u_2=4m+3$ only, is obtained:

$$\begin{aligned}&y_0(4m+1)=CHT[x(k)]=\\&=y_{0r}(4m+1)+j\,y_{0Im}(4m+1)=y_0^*(4m+3),\end{aligned} \quad (8)$$

$$M_0(4m+1)=M_0(4m+3)=\sqrt{y_{0r}(4m+1)^2+y_{0Im}(4m+1)^2}, \quad (9)$$

$$\varphi_0(4m+1)=-\varphi_0(4m+3)=arctg[y_{0Im}(4m+1)/y_{0r}(4m+1)], \quad (10)$$

$$\begin{aligned}&y_1^s(4m+1)=CHT[e_0(k)]=\\&=y_{1r}^s(4m+1)+j\,y_{1Im}^s(4m+1)=y_1^{s\,*}(4m+3)\end{aligned} \text{ for } s=1,2, \quad (11)$$

$$M_1^s(4m+1)=M_1^s(4m+3)=\sqrt{y_{1r}(4m+1)^2+y_{1Im}(4m+1)^2}, \quad (12)$$

$$\varphi_1^s(4m+1)=-\varphi_1^s(4m+3)=arctg[y_{1Im}^s(4m+1)/y_{1r}^s(4m+1)]. \quad (13)$$

The watermark elements $w_l(4m+1)$ and $w_l(4m+3)$, which are inserted in the SP levels l=0,1, modify the phases $\varphi_0(4m+1)$, $\varphi_0(4m+3)$ and $\varphi_1^s(4m+1)$, $\varphi_1^s(4m+3)$ for s=1,2 and l=0,1 of the corresponding coefficients $y_0(4m+1)$, $y_0(4m+3)$ and $y_1^s(4m+1)$, $y_1^s(4m+3)$ in accordance with the relations:

If $|M_l(4m+1)|>\eta$, then

$$\varphi_{lw}(4m+1)=\varphi_l(4m+1)+w_l(4m+1) \text{ and}$$
$$\varphi_{lw}(4m+3)=-\varphi_{lw}(4m+1) \quad (14)$$

Here η is a threshold, used to define the spectrum coefficients, suitable for watermarking. The value of η is selected to be the part α<1 of the module M_{max} of the biggest coefficient in the transform of the processed fragment; the minimum value of η should be higher than some small positive value δ, settled in advance, i.e.:

$$\eta=\begin{cases}\alpha M_{max} & \text{if } \alpha M_{max}>\delta,\\ \delta & \text{- in other cases.}\end{cases} \quad (15)$$

In the SP level l=1 for every value of s=1,2 the watermarking is performed, changing the corresponding phases $\varphi_1^s(4m+1)$ and $\varphi_1^s(4m+3)$ of the two couples of coefficients $y_1^s(4m+1)$ and $y_1^s(4m+3)$, i.e. there could be inserted two different values $w_1^1(4m+1)$ and $w_1^2(4m+1)$. In order to retain the subjective quality of the watermarked signal equal with that of the original one, the maximum value of each data element of the watermark w_l for a given SP level should be restricted in correspondence with the requirement $|w_l(u)|\leq 0{,}05$ rad ($\leq 3^0$). This ensures practical inaudibility of the changed phases of the spectrum components with frequencies (4m+1) and (4m+3) in the two SP levels, taking into account that the sequence of w_l elements must have a pseudorandom structure. The requirement for watermark data secrecy is suited, performing the function "exclusive OR" for each $w_l(u)$ element with its corresponding element from the pseudorandom sequence, which is the used secret key [5]. The watermarked audio signal $x_w^1(k)$, in accordance with (14) in the SP level l is:

$$x_w^0(k)=\tilde{x}_{0w}(k)+e_0(k)=\tilde{x}_{0w}(k)+x(k)-\tilde{x}_0(k) \quad \text{for } l=0, \qquad (16)$$

$$\begin{aligned} x_w^1(k)&=\tilde{x}_{0w}(k)+\tilde{e}_{0w}^s(k)+e_1(k)= \\ &=\tilde{x}_{0w}(k)+\tilde{e}_{0w}^s(k)+e_0(k)-\tilde{e}_0^s(k) \end{aligned} \quad \text{for } l=1 \text{ and } s=1,2 \qquad (17)$$

Here $x_w^0(k)$ contains the element w_0, and $x_w^1(k)$ - respectively the elements w_0, w_1^1 and w_1^2. The components $\tilde{x}_{0w}(k)$, $\tilde{e}_{0w}^s(k)$, $e_0(k)$, $e_1(k)$ in (16) and (17), are defined in correspondence with:

$$\tilde{x}_{0w}(k)=CHT^{-1}\{F_0[M_0(u)\exp\ {}_{0w}(u)]\}, \qquad (18)$$

$$e_0(k)=x(k)-CHT^{-1}\{F_0[M_0(u)\exp\varphi_0(u)]\}, \qquad (19)$$

$$\tilde{e}_{0w}^s(k)=CHT^{-1}\{F_1[M_1^s(u)\exp\ {}_{1w}^s(u)]\} \quad \text{for } s=1,2; \qquad (20)$$

$$e_1(k)=e_0(k)-CHT^{-1}\{F_1[M_1^s(u)\exp\varphi_1^s(u)]\}. \qquad (21)$$

Here, using $F_0[.]$ and $F_1[.]$ from levels l=0,1, are selected only the coefficients with frequencies (4m+1) and (4m+3), whose phases $\varphi_{0w}(4m+1)$ and $\varphi_{1w}^s(4m+3)$ for s=1,2 are modified in correspondence with Eq. (14). The described watermarking principle permits the insertion of (L+2R) watermark elements w_l in every N-dimensional audio fragment, and of 4R coefficients - in the level l=1.

2.2 Audio Watermark Extraction

The watermark extraction from the normalized watermarked audio signal $x_w^0(k)$ or $x_w^1(k)$, defined with Eqs. (16-17), is performed using the original audio signal x(k). The elements $w_0(4m+1)$ are presented as follows:

$$y_{0w}(u)=CHT\{x_w^0(k)\cup x_w^1(k)\}=M_0(u)\exp\varphi_{0w}(u),$$

$$y_0(u)=CHT\{x(k)\}=M_0(u)\exp\varphi_0(u) \quad \text{for } u=4m+1. \qquad (22)$$

Then, from Eqs. (14) and (21) for $\varphi_0(4m+1)\neq 0$, follows:

$$w_0(4m+1)=\varphi_{0w}(4m+1)-\varphi_0(4m+1). \qquad (23)$$

The elements $w_1^1(4m+1)$ and $w_1^2(4m+1)$ for s=1,2 are defined in accordance with:

$$y_{1w}^s(u)=CHT\{e_{0w}^s(k)\}=M_1^s(u)\exp\varphi_{1w}^s(u),$$

$$y_1^s(u)=CHT\{e_0^s(k)\}=M_1^s(u)\exp\varphi_1^s(u) \quad \text{for } u=4m+1. \qquad (24)$$

The differences $e_{0w}^s(k)$ and $e_0^s(k)$ for s=1,2 in accordance with (23) are equal to:

$$e_{0w}^s(k)=x_w^1(k)-\tilde{x}_{0w}(k),\ e_0^s(k)=x(k)-\tilde{x}_0(k) \text{ for } l=0,1 \qquad (25)$$

Here $x_w^1(k)$, $\tilde{x}_{0w}(k)$ and $\tilde{x}_0(k)$ are calculated, using Eqs. (17), (18) and (2).

In case, that $\varphi_1^s(4m+1)\neq 0$, the watermark elements for s=1,2 are defined with the relation:

$$w_1^s(4m+1)=\varphi_{1w}^s(4m+1)-\varphi_1^s(4m+1). \qquad (26)$$

2.3 Audio Watermark Detection

In order to ensure higher reliability of the watermark detection, the elements w_l, extracted from the normalized audio fragments, should be compared with their originals w_l^*. For this purpose is used the coefficient of the normalized cross-correlation ρ_l of the two sequences w_l and w_l^* [2]. The solution for the watermark detection in the first SP level is taken when the condition in (27) is answered:

$$\rho_l(w_l,w_l^*)=\left\{\sum_{n=1}^{N_l} w_l(n)w_l^*(n) \Big/ \sqrt{\sum_{n=1}^{N_l} w_l^2(n)}\sqrt{\sum_{n=1}^{N_l} w_l^{*2}(n)}\right\}\geq T_l \quad \text{for } l=0, \qquad (27)$$

Here T_l is a threshold, selected in advance, and N_l is the number of fragments, containing the w_l elements.

2.4 Algorithm for Watermarking of Audio Fragments

Based on the already described principle here follows the algorithm for watermarking of a couple of complex-conjugated CHT coefficients, generalized for M couples. In case, that from all the CHT coefficients of the N-dimensional audio fragment in the level l=0, we retain only the couple $y_0(4m+1)$ and $y_0(4m+3)$, the approximation model of the fragment is defined in accordance with Eqs. (2)-(6) and (8)-(9):

$$\begin{aligned}\tilde{x}_0(k)&=\left(\tfrac{1}{N}\right)[y_0(4m+1)j^{(4m+1)k}h(4m+1,k)+y_0(4m+3)j^{(4m+3)k}\times \qquad (28)\\ &\times h(4m+3,k)]=\left(\tfrac{1}{N}\right)M_0(4m+1)h(4m+1)j^k[e^{j\varphi_0(4m+1)}+(-1)^k e^{-j\varphi_0(4m+1)}].\end{aligned}$$

Here:

$$y_0(4m+1)=y_0^*(4m+3)=\sum_{k=0}^{N-1} x(k)h(4m+1,k)j^{-k}= =C_0(4m+1)+jD_0(4m+1) \quad (29)$$

$$C_0(4m+1)=\sum_{v=0}^{(N/4)-1}[x(4v)-x(4v+2)](-1)^{\alpha_0(m,v)},$$
$$D_0(4m+1)=\sum_{v=0}^{(N/4)-1}[x(4v+3)-x(4v+1)](-1)^{\alpha_0(m,v)}, \quad (30)$$
$$\alpha_0(m,v)=\sum_{r=3}^{\lg_2 N}\lfloor m/2^{r-3}\rfloor\lfloor v/2^{r-3}\rfloor.$$

The module and the phase of the coefficient $y_0(4m+1)$ are represented with:

$$M_0(4m+1)=\sqrt{C_0(4m+1)^2+D_0(4m+1)^2},$$
$$\varphi_0(4m+1)=\mathrm{arctg}[D_0(4m+1)/C_0(4m+1)]. \quad (31)$$

From Eqs. (28) and (31) follows that for k=4v,4v+1,4v+2,4v+3 and v=0,1,..,(N/4)-1 and in accordance with Eqs. (14) and (16) is obtained the model of the marked audio signal for the level l=0:

$$x_w^0(4v)=x(4v)-[\tilde{x}_0(4v)-\tilde{x}_{0w}(4v)]=x(4v)-a_0(m), \quad (32)$$

$$x_w^0(4v+1)=x(4v+1)-[\tilde{x}_0(4v+1)-\tilde{x}_{0w}(4v+1)]= = x(4v+1)-b_0(m), \quad (33)$$

$$x_w^0(4v+2)=x(4v+2)-[\tilde{x}_0(4v+2)-\tilde{x}_{0w}(4v+2)]= =x(4v+2)+a_0(m), \quad (34)$$

$$x_w^0(4v+3)=x(4v+3)-[\tilde{x}_0(4v+3)-\tilde{x}_{0w}(4v+3)]= =x(4v+3)+b_0(m), \quad (35)$$

where:

$$a_0(m)=\tfrac{2}{N}[C_0(4m+1)\beta_0(m)+D_0(4m+1)\delta_0(m)], \quad (36)$$

$$b_0(m)=\tfrac{2}{N}[C_0(4m+1)\delta_0(m)-D_0(4m+1)\beta_0(m)], \quad (37)$$

$\beta_0(m)=1-\cos[w_0(4m+1)]$ and $\delta_0(m)=\sin[w_0(4m+1)]$.

These relations are the basis of the algorithm for watermark insertion in the SP level l=0. At the beginning, in. (30) are calculated the values of $C_0(4m+1)$ and $D_0(4m+1)$, which are substituted in (36) and (37). Then from Eqs. (32)-(35) are calculated the values of the watermarked signal $x_w^0(k)$.

The watermark extraction from $x_w^0(k)$ is performed using the original x(k), $a_0(m)$ and $b_0(m)$, defined as:

$$a_0(m)=x(4v)-x_w^0(4v)=x_w^0(4v+2)-x(4v+2), \quad (38)$$

$$b_0(m)=x(4v+1)-x_w^0(4v+1)=x_w^0(4v+3)-x(4v+3). \quad (39)$$

From (30) are calculated $C_0(4m+1)$ and $D_0(4m+1)$, and are substituted in:

$$w_0(4m+1)= =\arcsin\left\{\frac{N[a_0(m)D_0(4m+1)+b_0(m)C_0(4m+1)]}{2[C_0(4m+1)^2+D_0(4m+1)^2]}\right\} \quad (40)$$

This is the solution of the system of Eqs. (36)-(37) concerning the element $w_0(4m+1)$ of the corresponding watermark for l=0. In similar way are inserted and extracted the watermarks w_1^1 and w_1^2 in the level l=1 for s=1,2. The described algorithm is generalized for the watermarking of M complex-conjugated couples of CHT coefficients in every SP level. Let M is integer, whose value is in the range from 1 to (N/4)-1. Then, in accordance with Eqs. (32)-(35) the watermarked signal for l=0 is:

$$x_w^0(4v)=x(4v)-A_0(M), \quad x_w^0(4v+1)=x(4v+1)-B_0(M), \quad (41)$$

$$x_w^0(4v+2)=x(4v+2)+A_0(M), \quad (42)$$

$x_w^0(4v+3)=x(4v+3)+B_0(M)$, where:

$$A_0(M)=\tfrac{2}{N}\sum_{m=0}^{M-1}[C_0(4m+1)\beta_0(m)+D_0(4m+1)\delta_0(m)] \quad (43)$$

$$B_0(M)=\tfrac{2}{N}\sum_{m=0}^{M-1}[C_0(4m+1)\delta_0(m)-D_0(4m+1)\beta_0(m)] \quad (44)$$

In these relations $C_0(4m+1)$ and $D_0(4m+1)$ are determined with (30). In the frequency band of the marked CHT coefficients from the SP level l=0 (with general number 2M) is reasonable to accept, that the watermark elements in the different couples are the same, i.e. $w_0(4m+1)=w_0$. This ensures better conditions for decreasing the watermark detection errors. In this case, taking into account the requirement $|w_0|\le 3^0$, from Eqs. (43),(44) and (30) follows:

$$A_0(M)\approx w_0(2/N)\sum_{m=0}^{M-1}D_0(4m+1)=w_0 S_1^0(M),$$
$$B_0(M)\approx w_0(2/N)\sum_{m=0}^{M-1}C_0(4m+1)=w_0 S_2^0(M), \quad (45)$$

where:

$$S_1^0(M)=(2/N)\sum_{m=0}^{M-1}\sum_{v=0}^{(N/4)-1}[x(4v+3)-x(4v+1)](-1)^{\alpha(m,v)}, \quad (46)$$

$$S_2^0(M)=(2/N)\sum_{m=0}^{M-1}\sum_{v=0}^{(N/4)-1}[x(4v)-x(4v+2)](-1)^{\alpha(m,v)}. \quad (47)$$

Then, for the extraction of the elements w_0 is used the relation:

$$w_0=A_0(M)/S_1^0(M)=B_0(M)/S_2^0(M), \quad (48)$$

$S_1^0(M)$ and $S_2^0(M)$ are defined from Eqs. (46)-(47), and $A_0(M)$ and $B_0(M)$ - from Eqs. (41) - (42).

$$A_0(M)=x(4v)-x_w^0(4v)=x_w^0(4v+2)-x(4v+2), \quad (49)$$

$$B_0(M)=x(4v+1)-x_w^0(4v+1)=x_w^0(4v+3)-x(4v+3). \quad (50)$$

The insertion/extraction of the watermarks w_1^1 and w_1^2 in the SP level l=1 for s=1,2 is performed in similar way. For the calculation of $C_1(4m+1)$ and $D_1(4m+1)$, respectively of $S_1^1(M)$ and $S_2^1(M)$, the N-dimensional audio signal x(k) is substituted with the difference $e_0^s(k)$, which contains N/2 discrete values.

The block diagram of the algorithm for multi-layer watermark embedding is presented in Fig.1. At the beginning is set the watermark data, which is rearranged in a pseudo-random sequence, using a public or secret key. The input audio signal is divided in fragments and each fragment is normalized. Then the data is prepared for the processing. All the following procedures are performed in accordance with the already described algorithm and are presented in the block diagram below.

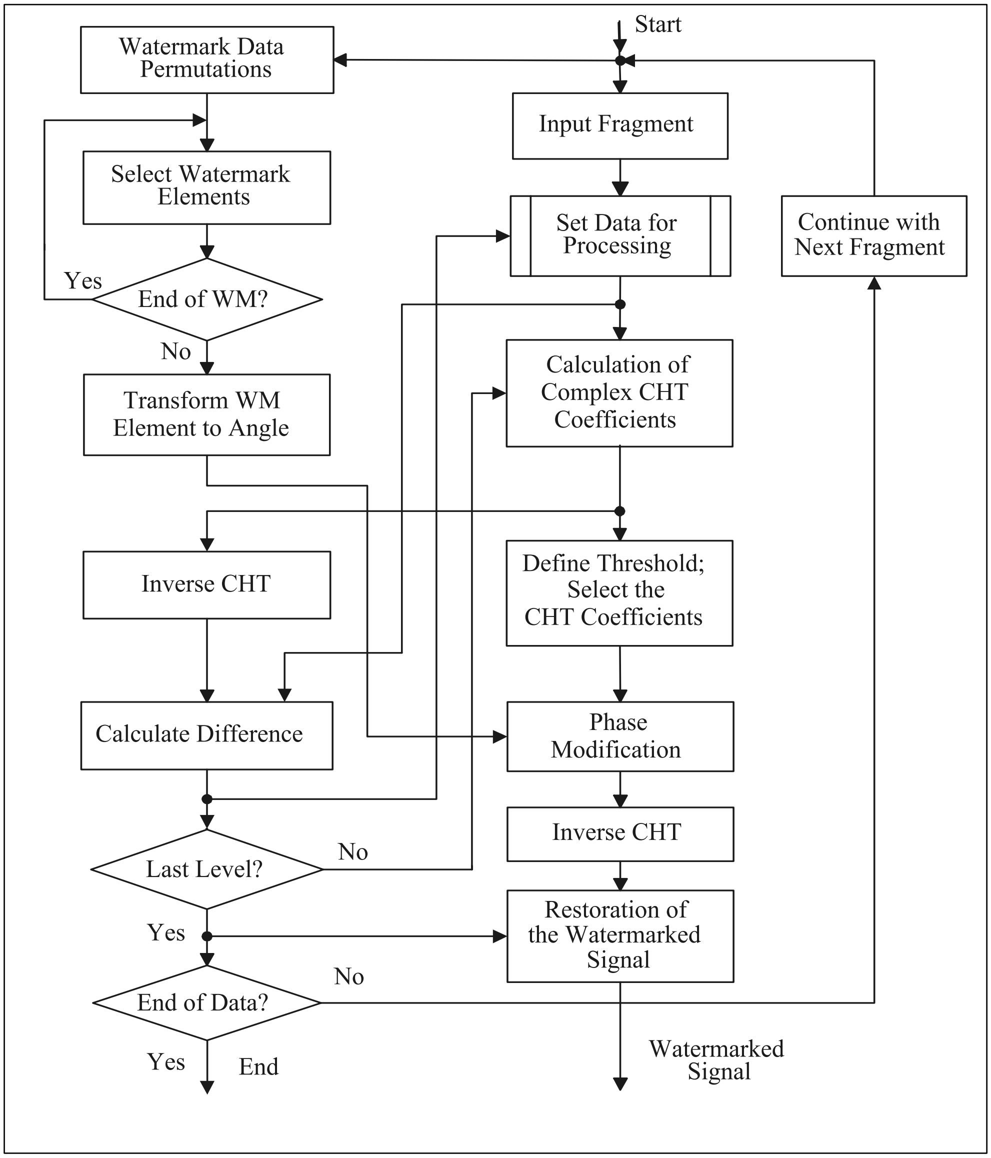

Fig.1 Block diagram of the algorithm for multi-layer watermark embedding

The block diagram of the algorithm for watermark extraction from the tested signal is given in Figure 2. The processing of the tested and the original signals starts after their initial synchronization and the separation of the fragment, which will be processed first. Then values of the fragment samples are normalized. The procedure continues in accordance with the already described algorithm.

In some cases the watermark extraction does not require the original audio signal. Instead, is used a correlation detection, based on the known sequence of watermark elements. The watermarked audio signal for the SP level l=0 could be represented in correspondence with Eqs. (40)-(41) and (44), with a sequence of N-dimensional vectors of the kind:

$$Z_0=X+G_0=X+w_0S_0(M)\ , \qquad (51)$$

where:

$X=[x(0),x(1),x(2),.,x(k),.,x(N-1)]^t$ is the vector of the original audio fragment;

$G_0=w_0S_0(M)$ - a vector with elements:

$g_0(k)=w_0s(k)$ for $k=0,1,..,N-1$;

Z_0 – a vector with elements $z_0(k)=x(k)+g_0(k)=x_w(k)$, representing the watermarked fragment

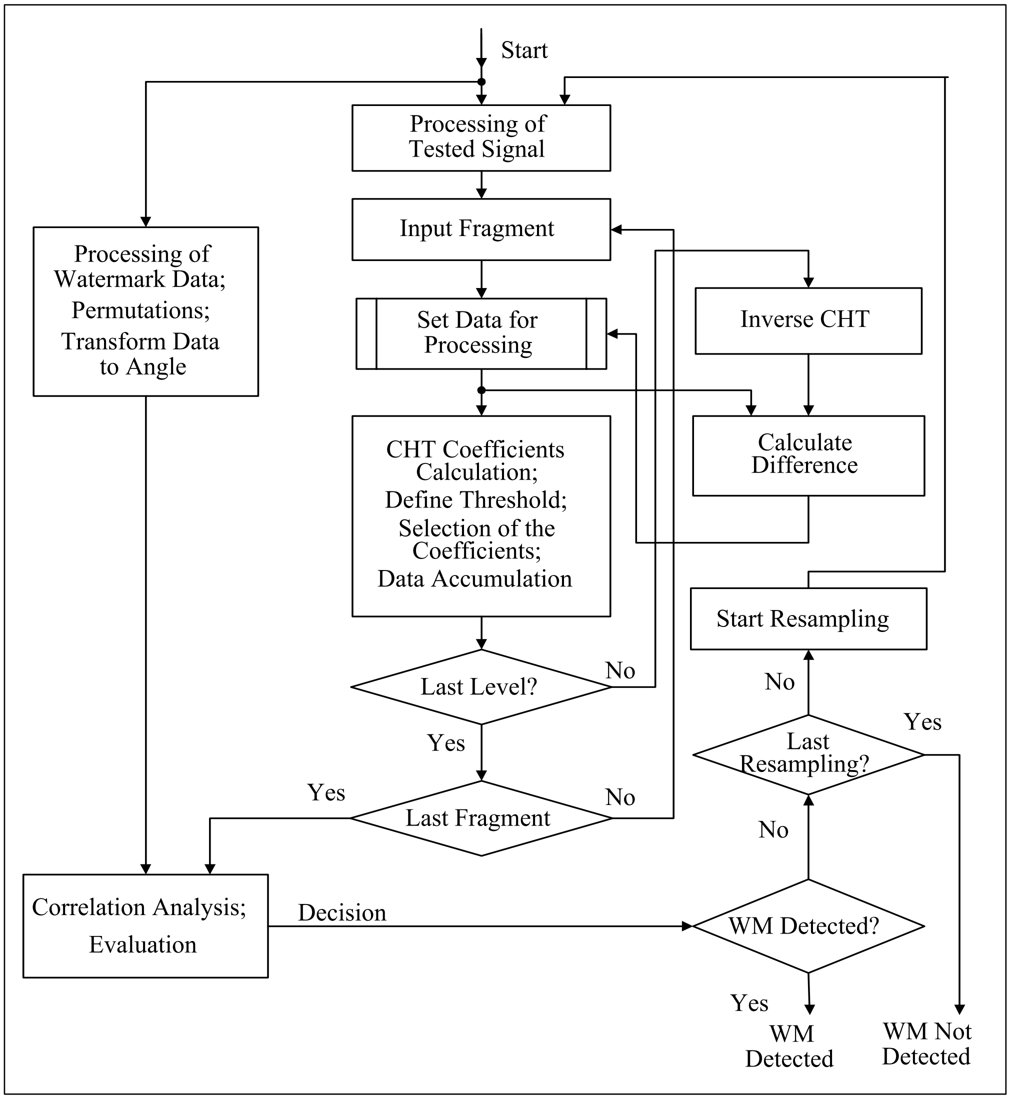

Fig.2 Block diagram for multi-layer watermark extraction

Let $W_0=w_0[-1,-1,+1,+1,...,-1,-1,+1,+1]^t$ represents the N-dimensional vector of the watermark in the SP level l=0. In this case, the coefficient of the normalized cross-correlation ρ for the couple of vectors Z_0 and W_0 from the sequence of N-dimensional audio fragments is defined with:

$$\rho_0(Z_0,W_0)=(Z_0\,W_0)/\|W_0\|^2= \\ =[1/(w_0N)]\left|\sum_{v=0}^{(N/4)-1}[x_w(4v+2)+x_w(4v+3)-x_w(4v)-x_w(4v+1)]\right| \quad (52)$$

The condition for the detection of the element w_0 of the watermark in the SP level l=0 for every audio fragment could be represented with the relation:

$$\rho_0(r)>T_0\ , \quad (53)$$

where:

$$\rho_0(r)=\left|\sum_{v=0}^{(N/4)-1}[x_w(4v+r+2)+x_w(4v+r+3)-x_w(4v+r)-x_w(4v+r+1)]\right|$$

For r=0,1,…, N-1 should be found the maximum of ρ_0. It must not be higher than the threshold T_0, which defines the probability for false alarm or missing the element w_0. The elements w_1^1 and w_1^2 in the SP level l=1 for every fragment could be extracted in similar way, using the corresponding correlation, ρ_1.

The block diagram of the algorithm for multi-layer watermark detection is presented in Figure 3. The process starts with the preparation of the tested signal and of the watermark data. Then consecutively are performed the permutation of the watermark data, using a public key, and its transform in corresponding angles.

The samples of the input fragment are normalized and are set for processing. After that the processing continues in accordance with the already described algorithm. At the end of the processing is taken a decision: WM detected / WM not detected and the search is finished. The main advantage of the described algorithm for watermark detection is its universality, because it does not require the original audio signal. In this case the probability for false detection is higher.

3. Evaluation of the Watermark Efficiency

As a criteria for the watermarking quality evaluation in every SP level could be used the mean square error (MSE) of the watermarked audio signal $x_w^0(k)$ or $x_w^1(k)$ in respect with the original, x(k). For a fragment of N discrete values MSE is defined with the relation:

$$\overline{\varepsilon_l^2} = \frac{1}{N}\sum_{k=0}^{N-1}[x(k)-x_w^l(k)]^2 \quad \text{for } l = 0,1. \tag{54}$$

Then, $SNR_l = 10\lg_{10}\{\sum_{k=0}^{N-1} x(k)^2 / \overline{\varepsilon_l^2}\}$ dB for l=0,1. (55)

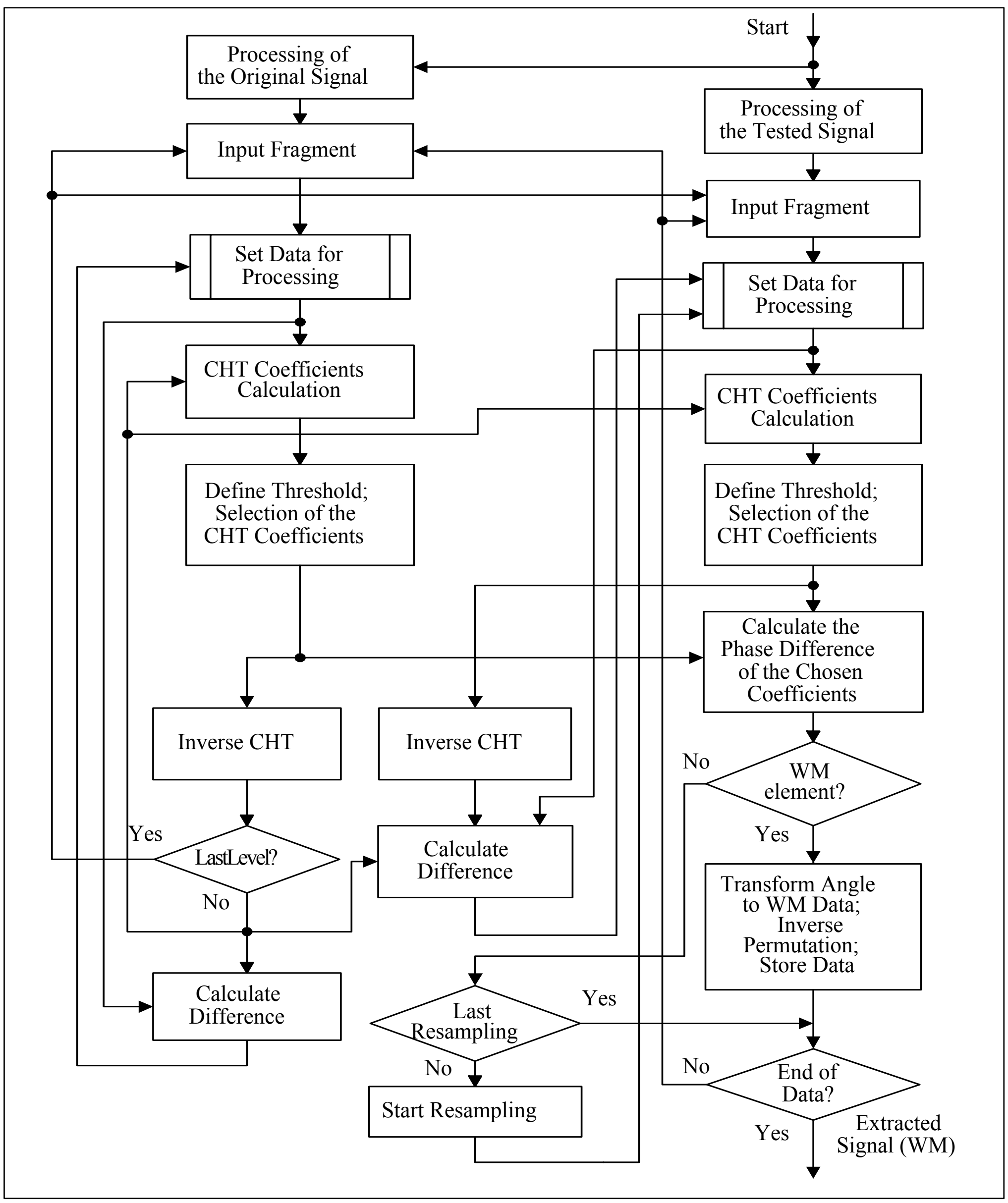

Fig.3 Block diagram of the algorithm for multi-layer watermarks identification

For l=0 from Eqs. (49), (50) and (54) is defined

$$\overline{\varepsilon_0^2}=[2(w_0)^2/N^2]\{[\sum_{m=0}^{M-1}C_0(4m+1)]^2+[\sum_{m=0}^{M-1}D_0(4m+1)]^2\},$$

from which for SNR_0 follows:

$$SNR_0= =10\lg\frac{N\sum_{k=0}^{N-1}x^2(k)}{2w_0^2\{[\sum_{m=0}^{M-1}C_0(4m+1)]^2+[\sum_{m=0}^{M-1}D_0(4m+1)]^2\}} \quad (56)$$

In similarity with (56) for the level l=1 and s=1,2:

$$SNR_1= =10\lg\frac{N\sum_{k=0}^{(N/2)-1}e^2(k)}{4w_1^2\{[\sum_{m=0}^{M-1}C_1(4m+1)]^2+[\sum_{m=0}^{M-1}D_1(4m+1)]^2\}} \quad (57)$$

From Eqs. (56-57) follows that SNR_l grows with the decreasing of the value of the watermark element w_l. The described method for audio watermarking was tested with great number of test audio signals with length 25 s, stored as WAVE files, with sampling frequency 44.1 kHz, 16 bits. The audio signals were divided in fragments, 256 samples each.

For L=R=1 and m=0 were modified the phases of the complex-conjugated couples for $y_0(1)$ and $y_0(3)$ from the level l=0, and $y_1^s(1)$ and $y_1^s(3)$ for s=1,2 - from the level SP l=1 for every fragment. The obtained results proved the high watermarking efficiency with SNR>75 dB, when the values of the watermark elements are in the range $\pm 3^0$ and are coded with 5 bits per sample. In this case the maximum speed for watermark data transmission in the SP level l=0 is approximately 860 bps, and in the SP level l=1 - correspondingly 1.72 Kbps.

Some results of the algorithms modeling with MATLAB 6.5 are presented in Figure 4. There are shown the original and the watermarked audio signal with 50 000 samples

In Figure 5 is shown the watermark with 5 bps, inserted in SP level l=0 for the left and right audio channel.

The watermark is inserted up to the 3[d] phase coefficient. The investigation on the algorithms regarding the basic pirate's attacks shows their high resistance, commensurable with that of methods, based on the wavelet transform [5,6].

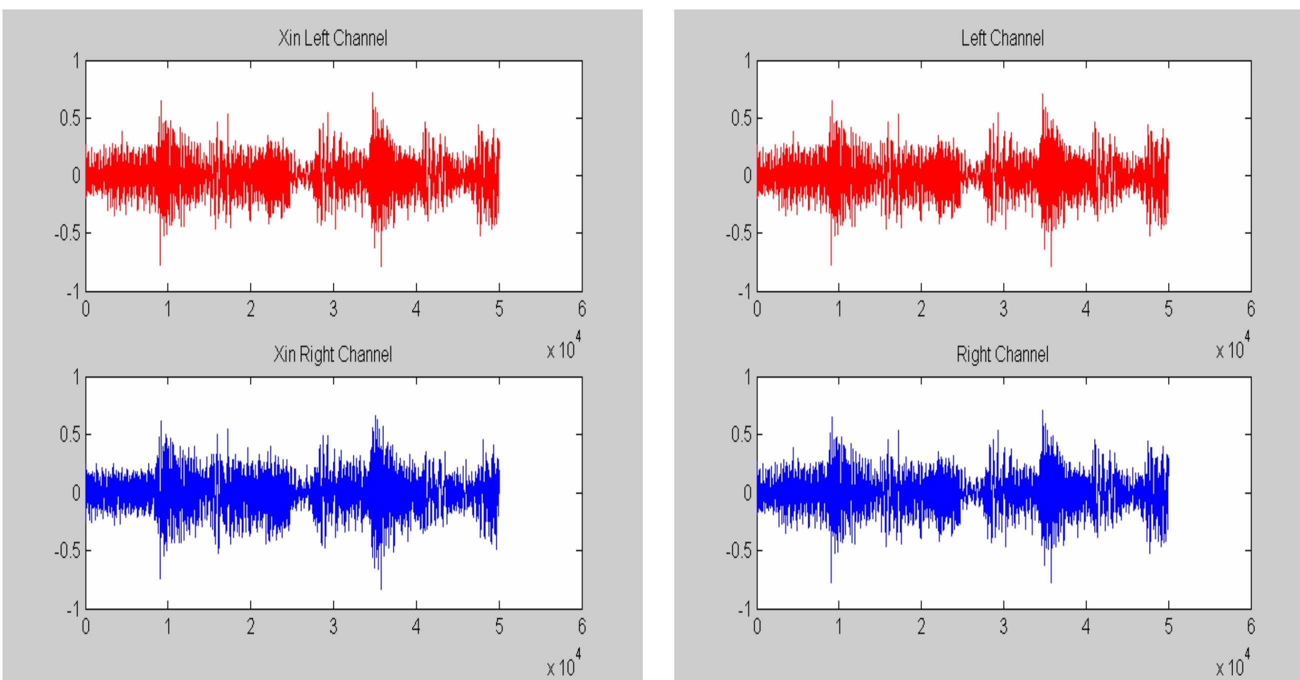

Fig.4. Original (left) and watermarked (right) audio signal with 50 000 samples

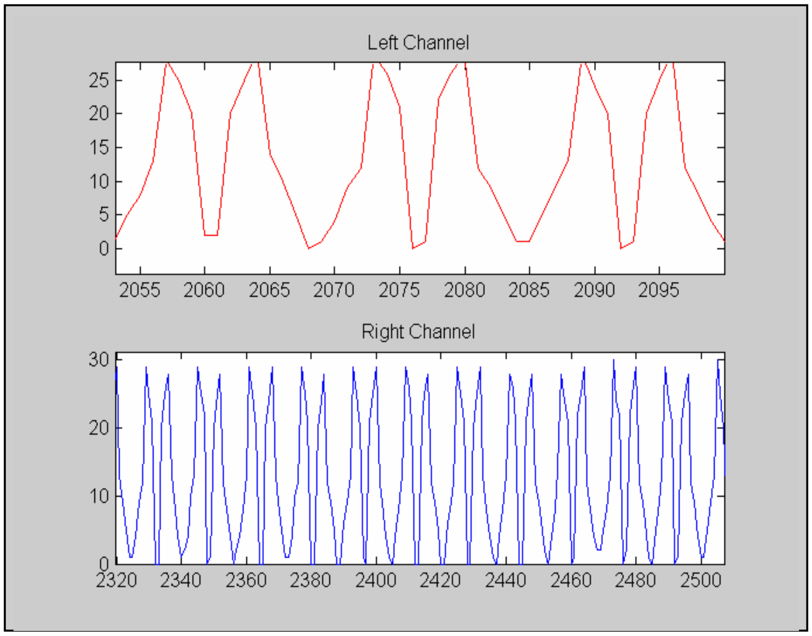

Fig.5. The watermark with 5 bps, inserted in level l=0 of SP for the left and the right audio channels.

4. Conclusion

A new method for audio watermarking in the phase spectrum with two-level SP was developed. The method is based on the limited spatial resolution of HAS in respect to the sound source direction, which results in practical inaudibility of the inserted watermarks. The method advantages are as follows: There is no quantization of the transform coefficients, causing additional noise in the watermarked signal; the method has relatively low computational complexity, which results from the fact that it is based on SP with CHT; the method has high resistance against pirates' attacks with multiple lossy compression or different kinds of audio transforms, because the watermark is embedded in the phase of the low-frequency spectrum coefficients and the phase modulation has higher immunity compared with the amplitude one; the method permits to insert different watermark with high information capacity in every SP level, which makes the identification of the manufacturers and of the authorized distributors of audio production much easier.

Acknowledgement

This paper was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science (Contract No. I-1304).

References

[1] M. Swanson B. Zhu, A.Tewfik, L. Boney. Robust audio watermarking using perceptual masking, Signal Processing, Vol. 66, 1998, pp. 337-355.
[2] I. Cox, M. Miller, J. Linnartz, T. Kalker. A Review of Watermarking Principles and Practices:Digital Signal Processing for Multimedia Systems, K. Parhi, T. Nishitani eds., Marcel Dekker, 1999, pp 461-482.
[3] A. J. Mason. Audio Watermarking - the State of the Art and Results of EBU Tests. BBC R&D White Paper, WHP 078, January 2004.
[4] A. Tewfik, Digital Watermarking. IEEE Signal Processing Magazine, Vol. 17, Sept. 2000, pp. 17-88.
[5] S. Katzenbeisser, F. Petitcolas eds. Information Hiding Techniques for Steganography and Digital Watermarking. Artech House, Boston 2000.
[6] C. Podilchuk, E. Delp. Digital watermarking: Algorithms and applications. IEEE Signal Processing Magazine, Vol. 14, No. 4, pp. 33-46, 2001.
[7] D. Kirovski, H. Malvar. Spread-Specrtum Watermarking of Audio Signals. IEEE Trans. on Signal Processing, Vol. 51, No. 4, April 2003, pp. 1020-1033.
[8] R. Garcia. Digital Watermarking of Audio Signals Using a Psycho-acoustic Auditory Model and Spread Spectrum Theory, 107-th AES Convention, Sept. 1999.

[9] C. Lu, H. Liao, L. Chen. Multipurpose Audio Watermarking, Proc. of the IEEE International Conference for Pattern Recognition'2000, pp. 282-285.
[10] H. Oh, J. Seok, J. Hong, D. Youn. New Echo Embedding Technique For Robust and Imperceptible Audio Watermarking, Proc. of ICASSP'2001, Vol. III, Salt Lake City, Utah, USA, May, 2001.
[11] P. Bassia, I. Pitas, N. Nikolaidis. Robust audio watermarking in the time domain. IEEE Trans. on Multimedia, vol. 3, No. 2, pp.232-241, 2001.
[12] B. Falkowski et al. Image Watermarking using the Complex Hadamard Transform, Proc. of the 33rd IEEE Symposium on Circuits and Systems, Geneve, May 2000, Vol. 4, pp. 573-576.
[13] R. Kountchev, M. Milanova, C. Ford, S. Rubin, R. Kountcheva. Audio Watermarking Based on Inverse Difference Decomposition, Proc. of the 2004 IEEE Intern. Conf. on Information Reuse and Integration (IRI-2004), Las Vegas, USA, November, 2004, pp. 180-185.

Roumen Kountchev received the M.Sc degree from Technical University of Sofia, Bulgaria in 1968. He received the Ph.D. degree from the Institute of Tele-communications, St. Petersbourg, Russia in 1975 and his D.Sc. degree from the Technical University of Sofia, Bulgaria in 2002. After working as an assistant professor (from 1968), and as an associate professor (from 1987) he has been a professor at the Technical University of Sofia, Bulgaria since 2003. His research interests include image processing, signal processing, image compression, multimedia watermarking. He is Member of the Technological Council of Bulgarian National Radio (from 2003), Member of the Higher Attestation Commission of the Council of Ministers of Bulgaria (from 2004) and President of BAPR (associated member of IAPR).

Dr. Mariofanna Milanova is Associate Professor of Computer Science Department, University of Arkansas at Little Rock, USA. She received the M. Sc. degree in Expert Systems and AI in 1991 and the PhD degree in Computer Science in 1995 from Technical University, Sofia, Bulgaria. She has extensive academic experience at various academic and research organizations including the Navy SPAWARS System Center, San Diego, USA, University of Louisville, USA, National Polytechnic Institute Research Center, Mexico, Technical University of Sofia, Bulgaria, University of Sao Paulo, Brazil, University of Porto, Portugal, Polytechnic University of Catalunya, Spain. She has been granted from the German Research Foundation, NATO and FAPESP State of Sao Paulo Research Foundation. Dr. Milanova is a senior member of the IEEE and Computer Society, member of IAPR, AAAS, member of the IEEE Women in Engineering and a member of the Cognitive Neuroscience Society. Milanova is editor of two books and several international conference proceedings. Her main research interests are in the areas of computer vision and communications, expert systems, artificial intelligence, data mining, multimedia watermarking, privacy and security based on biometric research.

Vladimir Todorov received the M.Sc degree from the Technical University of Sofia, Bulgaria in 1973. From 1973 to 1986 he was Head designer and Main designer in the Research Dept. of ZIT Plant, and from 1986 to 1992 - Research Director of the Technological Institute for Industrial Electronics (TIIE), Sofia. From 1993 he is the President of T&K Engineering, Sofia, Bulgaria**.**

Roumiana Kountcheva received the M.Sc degree in 1968 and the Ph.D. degree in 1977 from the Technical University of Sofia, Bulgaria. From 1973 to 1980 she was designer in the Research Dept. of ZIT Plant; from 1980 to 1992 she was senior researcher in the Technological Institute for Industrial Electronics (TIIE) and from 1986 to 1992 - Head of the Engineering Department in the same Institute. From 1993 she is the Vice-president of T&K Engineering, Sofia, Bulgaria.