Audio Watermarking in the Phase-Frequency Domain

R. Kountchev[1], R. Mironov[2]

Abstract - **A new approach for audio watermarking in the phase-frequency domain is presented, based on the Inverse Difference Decomposition with Complex Hadamard Transform. The main advantages are the absence of quantization noise, the lower computational complexity and the ability for the owner of the audio contents and for the authorized distributors to insert different watermarks in the protected audio signal.**

Keywords: digital audio watermarking, inverse difference decomposition, Complex Hadamard transform

I. INTRODUCTION

In correspondence with the up-to-date methods for audio watermarking [1,2], the watermarks are inserted in the time- or frequency domain of the audio signal, using some kinds of masking effects, which concern the sound perception in accordance with the human auditory system (HAS) [3]. In order to make the distortions, resulting from the watermark insertion in the time domain smaller [1], the watermark is presented as a pseudo-noise binary sequence, which is added to the corresponding discrete values of the audio segments. In the cases, when the watermarking is performed in the audio signal spectrum [2], usually are modified the amplitudes and the phases of selected complex low-frequency coefficients, obtained using one of the known discrete linear transforms: Fourier, Fourier-Mellin, Radon, etc. The second approach was developed for discrete cosine and wavelet transforms [4], for which the corresponding spectrum coefficients are real numbers. In this case is used modulation of selected cosine coefficients in the middle-frequency band of the audio signal, or the pseudo-random sequence is inserted in some components of its wavelet decomposition. The basic qualities of the spectrum approach for watermark insertion [1] are its practical inaudibility (transparency), the high resistance against pirates and fraud attacks, audio compression, time scaling, amplitude corrections, linear and nonlinear filtration, noising and applying of special effects.

In this work is offered a high-efficient method for audio watermarking, based on the modification of the phases of the spectrum coefficients in every level of their Inverse Difference Decomposition (IDD) with Complex Hadamard Transform (CHT), in correspondence with [5].

II. MATHEMATICAL DESCRIPTION

Audio Watermarking Based on IDD with CHT.

The new principle is based on CHT of the discrete audio signal components x(k) for k=0,1,..,N-1 and $N=2^n$, decomposed with two-level IDD, with number l=0,1:

$$x(k)=\tilde{x}_0(k)+\tilde{e}_0^s(k)+e_1(k)\,. \qquad (1)$$

For the level l=0 the component $\tilde{x}_0(k)$ is described with the approximation model of x(k), defined by the relations (2):

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

The operators CHT[•] and $CHT[\bullet]^{-1}$ represent the direct and the inverse CHT of the signal x(k), by (3):

$$y(u)=\sum_{k=0}^{N-1}x(k)t(u,k)=y_{Re}(u)+jy_{Im}(u)\,,\ x(k)=\frac{1}{N}\sum_{u=0}^{N-1}y(u)t^*(u,k)$$

where: $y_{Re}(u)$ and $y_{Im}(u)$ are correspondingly the real and the imaginary part of the complex coefficient $y(u)=M(u)\exp[j\varphi(u)]$, M(u) and $\varphi(u)$ represent the amplitude and the phase spectrums of the audio signal x(k), and t(u,k) and $t^*(u,k)$ are the elements of the matrices for the direct and inverse CHT:

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

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

Here $\lfloor * \rfloor$ is an operator, which represents the integer part of the result, obtained after the division.

For the level l=1 at first is defined the zero difference $e_0(k)=x(k)-\tilde{x}_0(k)$, from which is calculated its approximation model:

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

Here for s=1 the components $\tilde{e}_0^s(k)$ and $y_1^s(u)$ are defined from the difference $e_0(k)$, when k=0,1,..,(N/2)-1. For s=2 these components are defined from $e_0(k)$, when k=(N/2),(N/2)+1,..,N-1. The operators F_0[.] and F_1[.] represent the filtration of the spectrum of every segment in the levels l=0,1 of IDD, consisting in truncation of a selected part of the spectrum coefficients. The retained coefficients must be even number, complex-conjugated couples.

For the inverse CHT the "truncated" coefficients should be substituted with zeros. The last, residual component in the decomposition of Eq. (1) is defined as:

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

In particular, in case, that from the spectrum of every segment in the levels l=0,1 are selected the complex-

[1] Roumen Kountchev is with the Faculty of Telecommunications, Technical University of Sofia, Boul. Kl. Ohridsky, 8, Sofia 1000, Bulgaria. E-mail: rkountch@tu-sofia.bg

[2] Roumen Mironov is with the Fac. of Telecomm., TU of Sofia, Boul. Kl. Ohridsky 8, Sofia 1000, Bulgaria. E-mail: rpm@tu-sofia.bg

conjugated coefficients with frequencies $u_1=4m+1$ and $u_2=4m+3$ only, is obtained:

$$y_0(u_1)=\mathrm{CHT}[x(k)]=y_{0Re}(u_1)+j\,y_{0Im}(u_1)=y_0^*(u_2) \tag{8}$$

$$M_0(u_1)=M_0(u_2)\,, \qquad \varphi_0(u_1)=-\varphi_0(u_2)\,, \tag{9}$$

$$y_1^s(u_1)=\mathrm{CHT}[e_0(k)]=y_{1Re}^s(u_1)+j\,y_{1Im}^s(u_1)=y_1^{s*}(u_2)\,,\ s=1,2, \tag{10}$$

$$M_1^s(u_1)=M_1^s(u_2)\,,\ \ \varphi_1^s(u_1)=-\varphi_1^s(u_2) \tag{11}$$

The watermark elements w_l, which are inserted in the levels l=0,1 of IDD, modify the phases $\varphi_1^s(u_1)$ and $\varphi_1^s(u_2)$ of the corresponding spectrum coefficients $y_1^s(u_1)$ and $y_1^s(u_2)$ in accordance with:

$$\varphi_{lw}(u_1)=\varphi_l(u_1)+w_l(u_1)\,,\ \varphi_{lw}(u_2)=-\varphi_{lw}(u_1)\,. \tag{12}$$

In the level l=1 of IDD for every value of s=1,2 could be changed the corresponding phases $\varphi_1^s(u_1)$ and $\varphi_1^s(u_2)$ of the two coefficients couples $y_1^s(u_1)$ and $y_1^s(u_2)$, i.e. there could be inserted two different values $w_1^1(u_1)$ and $w_1^2(u_1)$. In order to retain the subjective quality of the watermarked signal equal with that of the original one, the maximum value of every element of the watermark w_l for a given IDD level must be restricted in correspondence with the requirement $|w_l|\leq 0{,}05$ rad ($\leq 3^0$). This ensures practical inaudibility of the changed phases of the spectrum components with frequencies (u_1) and (u_2) in the two IDD levels, in case that the sequence of w_l elements has a pseudorandom structure, without zero values. The requirement for watermark sign secrecy is answered performing the function “exclusive OR” for every w_l element with its corresponding element from the pseudorandom sequence, which represents the chosen secret key.

The audio signal $x_w^1(k)$, watermarked in accordance with Eq. (14) for the level l of IDD is correspondingly:

$$x_w^0(k)=\tilde{x}_{0w}(k)+e_0(k)=\tilde{x}_{0w}(k)+x(k)-\tilde{x}_0(k) \ \text{ for } l=0, \tag{13}$$

$$x_w^1(k)=\tilde{x}_{0w}(k)+\tilde{e}_{0w}^s(k)+e_0(k)-\tilde{e}_0^s(k) \ \text{ for } l=1 \text{ and } s=1,2 \tag{14}$$

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 $\tilde{x}_{0w}(k)$, $\tilde{e}_{0w}^s(k)$, in Eqs. (13) and (14), are defined in correspondence with:

$$\tilde{x}_{0w}(k)=\mathrm{CHT}^{-1}\{F_0[M_0(u)\exp\varphi_{0w}(u)]\}\,, \tag{15}$$

$$\tilde{e}_{0w}^s(k)=\mathrm{CHT}^{-1}\{F_1[M_1^s(u)\exp\varphi_{1w}^s(u)]\} \ \text{ for } s=1,2, \tag{16}$$

Here, using $F_0[\bullet]$ and $F_1[\bullet]$ from the spectrum in levels l=0,1, are selected only the coefficients with frequencies (u_1) and (u_2), whose phases $\varphi_{0w}(u_1)$ and $\varphi_{1w}^s(u_2)$ for s=1,2 are modified in correspondence with Eq. (14). The described watermarking principle permits the insertion of (L+2R) elements w_l of the watermark signs in every N-dimensional audio segment, modifying the phases of the 2L complex conjugated CHT coefficients from the level l=0, and of the 4R coefficients - in the level l=1 (L and R are selected in the interval from 1 to N/4).

The watermark extraction from the watermarked audio signal $x_w^0(k)$ or $x_w^1(k)$, defined from Eqs. (13) and (14), could be performed using the original signal x(k). In this case the elements $w_0(u_1)$ are defined using the relations:

$$y_{0w}(u_1)=\mathrm{CHT}\{x_w^0(k)\vee x_w^1(k)\}, \quad y_0(u_1)=\mathrm{CHT}\{x(k)\}\,. \tag{17}$$

Then, if $\varphi_0(u_1)\neq 0$, from Eqs. (14) and (21) follows:

$$w_0(u_1)=\varphi_{0w}(u_1)-\varphi_0(u_1)\,. \tag{18}$$

The elements $w_1^1(u_1)$ and $w_1^2(u_1)$ for s=1,2 are defined in accordance with:

$$y_{1w}^s(u_1)=\mathrm{CHT}\{e_{0w}^s(k)\}\,, \quad y_1^s(u_1)=\mathrm{CHT}\{e_0^s(k)\}\,. \tag{19}$$

In Eq. (19) the differences $e_{0w}^s(k)$ and $e_0^s(k)$ are equal to:

$$e_{0w}^s(k)=x_w^1(k)-\tilde{x}_{0w}(k)\,,\ e_0^s(k)=x(k)-\tilde{x}_0(k)\,. \tag{20}$$

Here $x_w^1(k)$, $\tilde{x}_{0w}(k)$ and $\tilde{x}_0(k)$ are calculated, using Eqs. (14), (15) and (2).

In case, that $\varphi_1^s(u_1)\neq 0$, the elements of the watermark are defined with the relation: $w_1^s(u_1)=\varphi_{1w}^s(u_1)-\varphi_1^s(u_1)$. (21)

In order to obtain higher reliability for the watermark detection, the elements w_l, extracted from the audio segments, must 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 IDD level is taken when the condition in Eq. (22) for l=0,1 is answered:

$$\rho_l(w_l,w_l^*)=\left\{\sum_{n=1}^{N_l} w_l(n)w_l^*(n) \Bigg/ \sqrt{\sum_{n=1}^{N_l} w_l^2(n)}\sqrt{\sum_{n=1}^{N_l} w_l^{*2}(n)}\right\}\geq T_l$$

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

Algorithm for Watermarking of Audio Segments.

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 segment in the level l=0, we retain only the couple $y_0(u_1)$ and $y_0(u_2)$, its approximation model is defined in accordance with the Eqs. (2)-(6) and (8)-(9), as follows:

$$\begin{aligned}\tilde{x}_0(k)&=(1/N)[y_0(u_1)j^{(u_1)k}h(u_1,k)+y_0(u_2)j^{(4m+3)k}h(u_2,k)]=\\&=(1/N)M_0(u_1)h(u_1)j^k[e^{j\varphi_0(u_1)}+(-1)^k e^{-j\varphi_0(u_1)}].\end{aligned} \tag{23}$$

Here:

$$y_0(u_1)=y_0^*(u_2)=\sum_{k=0}^{N-1} x(k)h(u_1 1,k)j^{-k}=C_0(u_1)+jD_0(u_1) \tag{24}$$

$$C_0(u_1)=\sum_{v=0}^{(N/4)-1}[x(4v)-x(4v+2)](-1)^{\alpha_0(m,v)},$$
$$D_0(u_1)=\sum_{v=0}^{(N/4)-1}[x(4v+3)-x(4v+1)](-1)^{\alpha_0(m,v)},\qquad(25)$$

where: $\alpha_0(m,v)=\sum_{r=3}^{\lg_2 N}\lfloor m/2^{r-3}\rfloor\lfloor v/2^{r-3}\rfloor$.

The module and the phase of the coefficient $y_0(u_1)$ are represented with:

$$M_0(u_1)=\sqrt{C_0(u_1)^2+D_0(u_1)^2},\ \varphi_0(u_1)=\mathrm{arctg}[D_0(u_1)/C_0(u_1)].\qquad(26)$$

From Eqs. (27-30) follows that after watermarking in accordance with the Eqs. (12-13) is obtained the marked audio signal for the level l=0 - $x_w^0(p)=x(p)-[\tilde{x}_0(p)-\tilde{x}_{0w}(p)]$. For p=4v, 4v+1,4v+2,4v+3 and v=0,1,..,(N/4)-1 following:

$$x_w^0(4v)=x(4v)-a_0(m),\qquad x_w^0(4v+1)=x(4v+1)-b_0(m),\qquad(27)$$
$$x_w^0(4v+2)=x(4v+2)+a_0(m),\qquad x_w^0(4v+3)=x(4v+3)+b_0(m),\qquad(28)$$

where: $a_0(m)=\frac{2}{N}[C_0(u_1)\beta_0(m)+D_0(u_1)\delta_0(m)],\qquad(29)$

$b_0(m)=\frac{2}{N}[C_0(u_1)\delta_0(m)-D_0(u_1)\beta_0(m)],\qquad(30)$

$\beta_0(m)=1-\cos[w_0(u_1)$, $\delta_0(m)=\sin[w_0(u_1)$,

These relations are used as a base for the algorithm for watermark insertion in the IDD level l=0. At the beginning, in the Eq. (25) are calculated the values of $C_0(u_1)$ and $D_0(u_1)$, which after that are substituted in Eqs. (29) and (30) for the calculation of $a_0(m)$ and $b_0(m)$. Then, from Eqs. (27)-(28) are calculated the values of the marked 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 with the differences:

$$a_0(m)=x(4v)-x_w^0(4v)=x_w^0(4v+2)-x(4v+2),\qquad(31)$$
$$b_0(m)=x(4v+1)-x_w^0(4v+1)=x_w^0(4v+3)-x(4v+3).\qquad(32)$$

From Eq. (25) are calculated $C_0(u_1)$ and $D_0(u_1)$, and together with the obtained values for $a_0(m)$ and $b_0(m)$ they are substituted in:

$$w_0(u_1)=\arcsin\left\{\frac{N[a_0(m)D_0(u_1)+b_0(m)C_0(u_1)]}{2[C_0(u_1)^2+D_0(u_1)^2]}\right\}\qquad(33)$$

The last relation is the solution of the system of Eqs. (29)-(30) concerning the element $w_0(u_1)$ of the corresponding watermark in the case, when l=0. In similar way are inserted and extracted the watermarks w_1^1 and w_1^2. The described algorithm is generalized for the watermarking of M complex-conjugated couples of CHT coefficients in every level of IDD.

In the cases, when the watermark extraction does not require the original audio signal, could be used correlation detection based on the sequence of watermark elements, known in advance:

- In correspondence with Eqs. (40)-(45), for the level l=0 of the IDD the watermarked audio signal could be represented with a sequence of N-dimensional vectors of the kind: $Z_0=X+G_0=X+w_0S_0(M)$, where: $X=[x(0),\ x(1),\ x(2),...,\ x(k),...,\ x(N-1)]^t$ is the vector of the original audio segment; $G_0=w_0S_0(M)$ - vector with elements $g_0(k)=w_0s(k)$ for $k=\overline{0,N-1}$ and Z_0 – vector with elements $z_0(k)=x(k)+g_0(k)=x_w(k)$, corresponding with the watermarked audio segment.

- Let $W_0=w_0[-1,-1,+1,+1,...,-1,-1,+1,+1]^t$ represents the N-dimensional vector for watermark in the level l=0. In this case, the coefficient of the normalized cross correlation ρ_0 of the couple of vectors Z_0 and W_0 from the sequence of N-dimensional audio segments is defined with:

$$\rho_0(Z_0,W_0)=(Z_0\,W_0)/\|W_0\|^2=$$
$$=[1/(w_0N)]\left|\sum_{v=0}^{(N/4)-1}[x_w(4v+2)+x_w(4v+3)-x_w(4v)-x_w(4v+1)]\right|\qquad(34)$$

The condition for the detection of the element w_0 of the watermark in the level l=0 of IDD for every audio segment could be represented with the relation: $\rho_0(r)>T_0$, 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 bigger 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 level l=1 of IDD for every audio segment could be extracted in similar way, using the corresponding correlation, ρ_1.

The main advantage of the described algorithm for watermark extraction is its universality, due to the fact, that it does not require the use of the original audio signal. In this case, however, the probability for false detection or missing is higher.

III. Evaluation of the Watermark Efficiency

As criteria for the watermark quality evaluation in every IDD 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 to the original, x(k). For a segment of N discrete values MSE is defined from the relation:

$$\overline{\varepsilon_l^2}=\frac{1}{N}\sum_{k=0}^{N-1}[x(k)-x_w^l(k)]^2\quad\text{for } l=0,1.\qquad(35)$$

Then the signal/noise ratio (SNR) is defined with:

$$SNR_l=10\lg_{10}\{\sum_{k=0}^{N-1}x(k)^2/\overline{\varepsilon_l^2}\}\ \text{dB}\quad\text{for } l=0,1.\qquad(36)$$

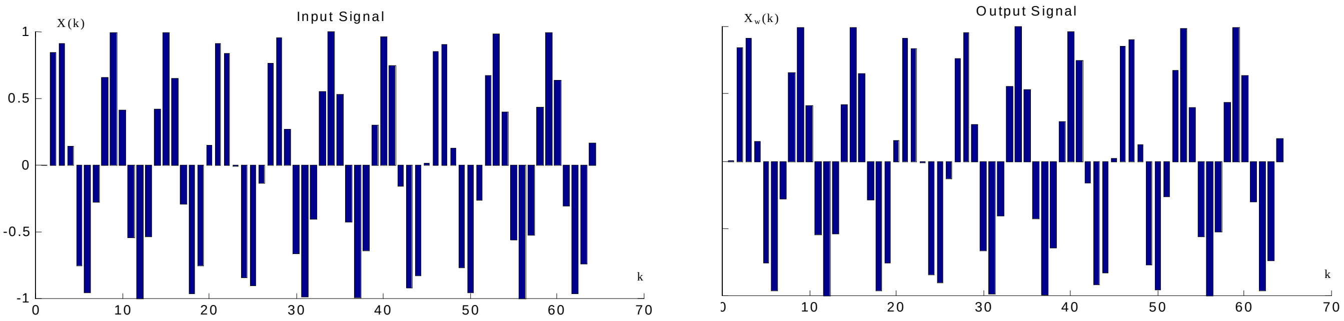

Fig.1. The original and watermarked test signals with 64 samples each.

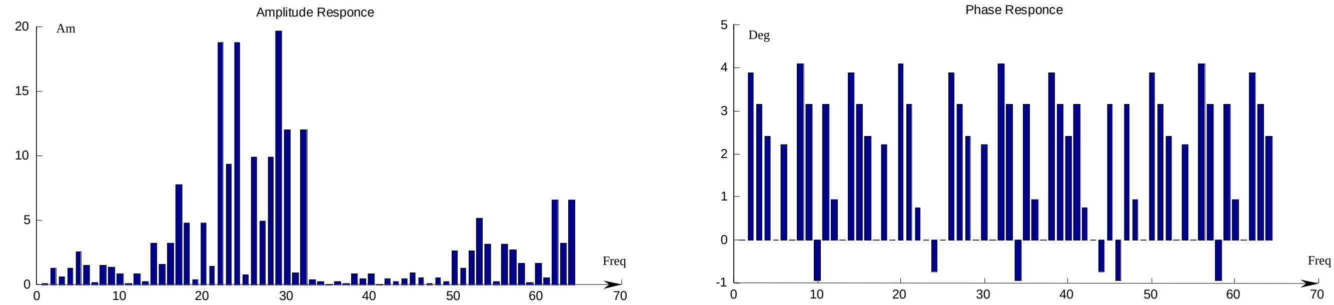

Fig.2. Amplitude and phase responses of the tested signal.

From the Eqs. (35) and (36) follows that the relation SNR_l grows together with the increasing of N and 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 25s, stored as WAVE files, with sampling frequency 44.1 kHz, 16 bits. The audio signals were divided in segments, 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 l=1 of IDD for every segment. The obtained results proved the high efficiency of the watermarking with SNR>80 dB when the values of the watermark elements are in the range $\pm 3^0$ and are coded with 5 bits per element, missing the code 00000. In this case the maximum speed for watermark data transmission in the level l=0 is approximately 860 bps, and in the level l=1 - correspondingly 1.72 кbps.

On Fig. 1 an input test sinusuidal signal with 64 samples, and the corresponding marked signal are presented. The 5 bits watermark is set to the 3th coefficient. The amplitude and phase responses of the input signal are presented on Fig.2. These results were obtained with program simulitaion on MATLAB 5.3.

IV. CONCLUSION

A new method for audio watermarking in the phase spectrum with two-level IDD, was developed. The method is based on the limited spatial resolution of the Human Auditory System in respect to the direction of the sound source, which results in practical inaudibility of the inserted watermarks. Advantage of the method is the fact that there is no quantization of the transform coefficients, the method has relatively low computational complexity and permits the insertion of different watermark with high information capacity in every IDD level.

The method offers exact watermark extraction, high resistance against frauds and pirates' attacks with multiple lossy compression or different kinds of audio transforms and the ability for exact watermark extraction without using the original audio signal. The insertion of different watermarks in every IDD level makes the identification of the manufacturers and of the authorized distributors of multimedia production much easier.

ACKNOWLEDGEMENT

The authors thank the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science for the financial support by the contract I-1304/2003.

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] F. Hartung, M. Kutter. Multimedia Watermarking Techniques. Proc. IEEE, Vol. 87, No 7, 1999, pp. 1079-1107.

[3] R. Garcia. Digital Watermarking of Audio Signals Using a Psycho-acoustic Auditory Model and Spread Spectrum Theory, 107-th AES Convention, Sept. 1999.

[4] X. Wang, H. Yang. A Content-Based Fragile Digital Audio Watermarking Algorithm. IEEE Intern. Conf. on Information Reuse and Integration, IRI'2003, Las Vegas, USA, October, 2003.

[5] R. Kountchev, M. Milanova, C. Ford, S. Rubin. Multimedia Watermarking with Complex Hadamard Transform in the Inverse Pyramid Decomposition, IEEE Intern. Conf. on Information Reuse and Integration, IRI'2003, Las Vegas, USA, October, 2003, pp. 305-310.