Real-Time MCLT Audio Watermarking and Comparison of Several Whitening Methods in Receptor Side

Marcado de Agua de Señales de Audio en Tiempo Real Usando MCLT y la Comparación de Varios Métodos para Decorrelacionar la Señal en el Receptor

José García Hernández, Mariko Nakano Miyatake and Héctor Pérez Meana
Escuela Superior de Ingeniería Mecánica y Eléctrica Unidad Culhuacán
Instituto Politécnico Nacional, Av. Santa Ana 1000, 04430 México D.F.
hmpm@prodigy.net.mx; mariko@calmecac.esimecu.ipn.mx

Article received on September 20, 2006; accepted on September 27, 2007

Abstract
A Real-Time audio watermarking scheme is proposed, where the strength of audio signal modifications is limited by the requirement of producing an output audio signal that is perceptually equal to the original one. The watermark embedding stage, based on a spread spectrum algorithm operating in the Modulated Complex Lapped Transform (MCLT) domain, inserts a watermark that is generated using a private key which modeled according to the Human Auditory System (HAS). The proposed blind detection approach is based on the Additive White Gaussian Noise (AWGN) channel theory; and to achieve this goal several whitening methods in the receptor side were evaluated. Evaluation results show that the proposed watermarking embedding scheme is robust to common attacks such like, D/A and A/D conversion, filtering, additive noise and high quality MPEG audio coding.
Keywords: Audio watermarking, copyright protection, Modulated Complex Lapped Transform, DSP implementation, whitening methods.

Resumen
Se propone un esquema para marcado de agua en tiempo real, donde la modificación introducida a las señales de audio esta limitada por la necesidad de producir una señal de salida que sea perceptualmente igual a la salida original. El sistema propuesto usa una detección a ciegas. Para llevar acabo la inserción se emplea un esquema de dispersión del espectro basado en la transformada MCLT. Aquí la marca de agua se genera usando una llave privada y modelado de acuerdo al sistema auditivo humano. Se lleva a cabo además una comparación de diversos métodos que pueden ser empleados para decorrelacionar la señal marcada en el receptor. La marca de agua insertada es robusta contra ataques comunes tales como: conversión A/D y D/A, filtrado, adición de ruido y codificación MPEG entre otros.
Palabras clave: Marcado de agua para señales de audio, protección de derecho de autor, Transformada MCLT, realización en DSP, métodos para decorrelacionar.

1 Introduction

The fast growing of internet services has increased the reproduction and retransmission of multimedia contents and as a consequence, both legal and unauthorized data manipulation has also grown. To avoid this problem, several methods have been proposed during the last decade. Among them a suitable solution to this problem is the watermarking approach, in which an imperceptible and statistically undetectable signature is added to the multimedia content; which to be useful for copyright protection, the watermark must completely characterize the person who embedded it and any unauthorized removing or manipulation of the watermark must render the audio material useless. The watermark also should satisfy another set of requirements such as: the watermark must be statistically undetectable and imperceptible to preserve the original signal quality; as well as it must be robust to signal processing operations such as: filtering, compression, resampling and noise contamination, etc.

A watermarking system consists, mainly, of three stages: a) the watermark generator, b) the watermark embedding stage and c) the watermark detector. Regarding this topics several algorithms for watermark embedding and watermark detection in audio sequences have bee proposed in the literature (Bassia *et al.*, 2001; Swanson *et al*, 1998; Kirovski *et al*, 2001; Neubauer *et al* 1998; Haitsma *et al*, 2000; Wender *et al*, 1996; Kirovski *et al*, 2003), that take advantage of the perceptual properties of the human auditory system (HAS); as well as from the MPEG audio coding standard. Kirovski and Malvar (2003) proposed a robust audio watermarking system in the MCLT domain in which only the coefficients between 200 Hz and 2000 Hz are marked; while only the audible coefficients are considered in the detection stage. This system spreads the watermark signal along several MCLT blocks and as a result, a typical implementation of it has a bit rate of 0.5-1 b/s. To increase the bit rate, this paper proposes a watermark embedding algorithm in which the watermark is spread along all frequency range of each MCLT block, such that the proposed system achieves a bit rate of about 11 b/s. To achieve an inaudible watermarking, this paper propose to model the watermark signal according to the HAS, previously used by Cvejic *et al*, 2001; Garcia *et al*, 2005. We also report the DSP implementation in a state-of-the-art Digital Signal Processor.

It is well known that when a spread spectrum method is used in an AWGN communications system the optimal detector is based on the correlation operation (Feller, 1968). However, because the magnitudes of MCLT coefficients do not present a Gaussian distribution, it is necessary to whiten them in order to use a correlation based watermark detector. To determine a suitable whitening method, we carried out a comparison among five different whitening methods, three of them previously reported (Kirovski *et al*, 2003; Garcia *et al*, 2005; Cvejic *et al*, 2001), such as the Cepstrum, LPC and Savitzky-Golay filtering methods; as well as the Median filter and Mean filters which, despite they are widely use in image signal processing operations, they have not been used in the digital watermarking fields. The experimental results show that the whitening methods based in Median and Mean filters are more reliable than the previously reported filters.

This paper is organized as follows: Section 2 presents the watermarking embedding process and the considerations for real-time implementation. The detection module and comparison of whitening methods are presented in Section 3, a statistical analysis of the detector is also shown. The evaluation results about the imperceptibility, the effectiveness of whitening procedures and robustness of the proposed system are presented in Section 4. Finally the conclusions are given in Section 5.

2 Watermarking Embedding

The watermarking embedding scheme, shown in figure 1, modifies the original signal using a spread spectrum approach; in which the original audio signal is represented as a 16-bit mono aural sequence sampled at 44100 Hz. The Pseudorandom noise, PN-sequence, is a pseudorandom clip frame with a Gaussian probability density function with zero mean and unit variance, pdf N(0,1), (Cvejic *et al*,2001); the PN-sequence length is 2048 and Alpha is a robustness factor.

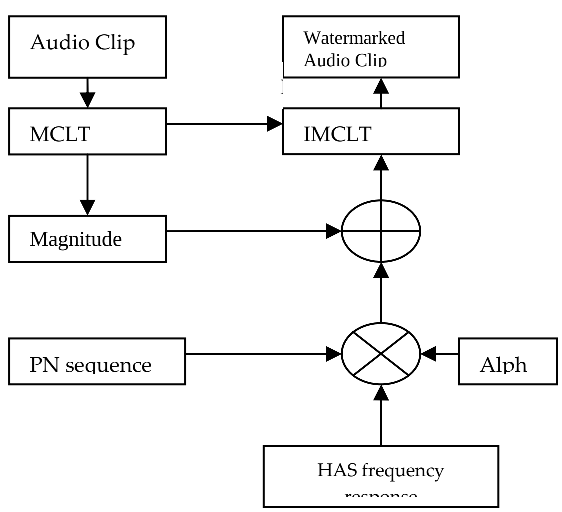


Fig. 1. Watermark embedding scheme

According to the block diagram shown in figure 1, the operation of the embedding stage is as follows: Firstly, the audio signal is divided in frames 4096 samples length, with 50% overlap. Next each block is transformed by using the MCLT. The magnitude and phase of MCLT coefficients are computed. Subsequently a given PN-sequence is frequency filtered by multiplying it by the HAS (human auditory system) in quiet conditions (Garcia et al, 2005), where the threshold of the HAS curve in quit conditions is shown in figure 2. The resulting curve is then multiplied by a strength factor called Alpha as shown in figure 1. Finally the watermarked signal is obtained by adding modified PN-sequence to the magnitude of original MCLT coefficients, keeping the phase without change, and applying inverse MCLT transform to them. Next subsections describe each one of the stages required for watermark embedding.

2.1 Watermark generator

To generate the watermark a noise generator, based on a Linear Feedback Shift Register (LFSR) developed from the primitive polynomial $x^{128}+x^{95}+x^{57}+x^{45}+x^{38}+x^{36}+1$ is used. It provides a period equal to 2^{128} bits; which is long enough for watermarking applications. Subsequently this pseudorandom sequence, which is assumed to be in the frequency domain is processed, in the frequency domain, to improve the watermark embedding using most of the HAS properties. The basic idea is that the pseudorandom sequence be shaped in accordance to the HAS in order to make the watermark even more imperceptible, by adjusting it the masking thresholds of the HAS in the frequency domain. This allows adapting the watermark in such way that its energy is maximized while keeping the auditory distortions to a minimum, although the signal-to-noise ratio (SNR) value becomes significantly decreased. The frequency characteristics of the filter are the approximation of the threshold in quiet curve of HAS, shown on figure 2. Finally the filtered sequence is multiplied by a constant Alpha, that represents a trade-off between the perceptual transparency and detection reliability. A suitable value of Alpha is between 0.5 and 1.5 dB because in this range, well trained hears are not able to distinguish between the original and watermarked audio signals (Kirovski, and Malvar, 2003).

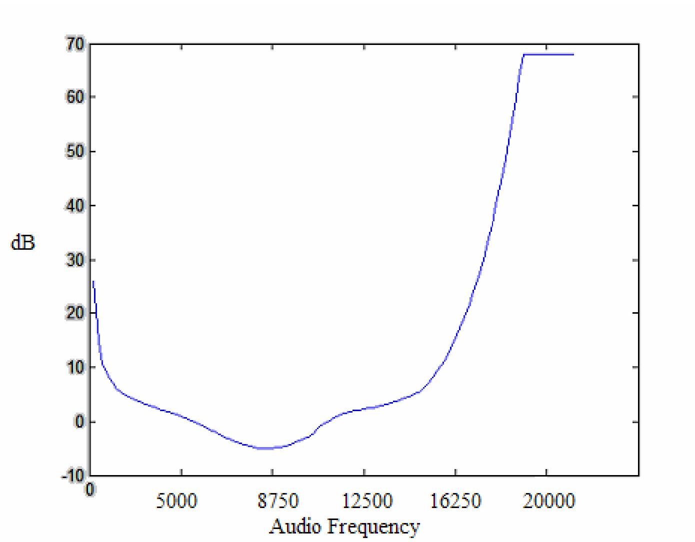


Fig. 2. Threshold in quiet curve of HAS

Another parameter that must be determined is the block size N. For this purpose several tests were carried out in which the correlation value obtained of the detected watermark is estimated, which are shown in figure 3. From this figure is possible to see that from N=512 to N=2048 there is an important improvement as N grows; however as N increases the computational complexity also increase and for values of N larger than 2048 the increase of the correlation value is not significant taking in account the increase on the computational complexity. Thus a value of

N equal to 2048 was selected because it provides a good tread off between computational complexity, accurate detection and embedding capacity.

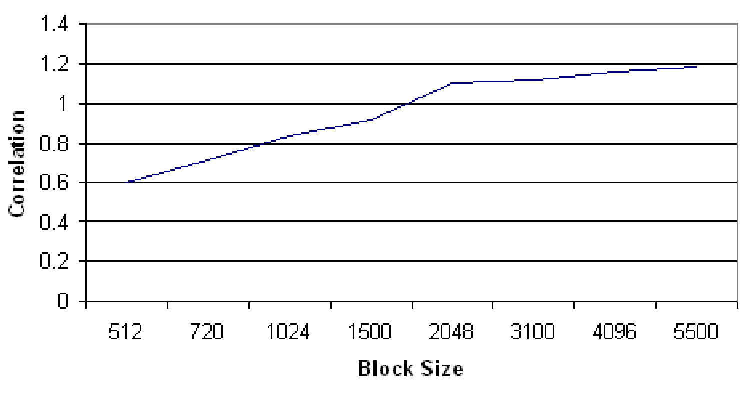


Fig. 3. Relation between block size and correlation value in proposed watermarking system

2.2 MCLT Modulated complex lapped transform

The MCLT is a particular kind of a oversampled generalized DFT filter bank proposed by Malvar (1999) which is given for analysis as:

$$p_a(n,k) = p_a^c(n,k) - jp_a^s(n,k) \tag{1}$$

$$p_a^c(n,k) = h_a(n)\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \tag{2}$$

$$p_a{}^s(n,k) \equiv h_a(n)\sqrt{\frac{2}{M}}\sin\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \tag{3}$$

and synthesis:

$$p_s(n,k) = \frac{1}{2}\left[p_s^c(n,k) - jp_s^s(n,k)\right] \tag{4}$$

$$p_s^c(n,k) \equiv h_s(n)\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \tag{5}$$

$$p_s{}^s(n,k) \equiv h_s(n)\sqrt{\frac{2}{M}}\sin\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right] \tag{6}$$

With

$$h_a(n) = h_s(n) = -\sin\left[\left(n+\frac{1}{2}\right)\frac{\pi}{2M}\right] \tag{7}$$

n and k are the time and frequency indexes, respectively. Thus from equations (1)-(7) it follows that the MCLT coefficients of a given input vector **x** an be estimated as $\mathbf{X} = \mathbf{P}_a^T \mathbf{x}$, while the Inverse MCLT of a vector **X** is given by $\mathbf{y} = \mathbf{P_s X}$. One of the most interesting MCLT properties is the fact that it is possible to recover original signal from MCLT coefficients by using only real part of MCLT coefficients, only its imaginary part or both of them. To reduce the processing requirements in a Real-time implementation, in the proposed system, only the real part of MCLT coefficients will be used (Malvar, 1999).

3 Watermark Detection

Several assumptions must be done to develop a reliable watermarking detection module able of providing the desired performance and robustness. Firstly, the watermark should be detected after some simple signal manipulations; next in most watermarking applications the original signal is not available in the detection stage such that a blind detector is required. Finally we assumed that the watermarking scheme can be considered as an AWGN communications system, where the audio signal is the communications channel and the watermark is the massage. Thus, under these assumptions, from detection theory it follows that the optimal blind detector becomes the correlation scheme (Feller, 1968). However the magnitude of the MCLT of audio signals are not a white Gaussian sequence (Kirovski *et al*, 2003), because the audio samples are highly correlated; and then a whitening procedure is required. Thus to develop a robust blind watermarking detection method based on a correlation approach, several whitening methods proposed during the last several years will be analyzed such as the LPC, Cepstrum filtering and Savitzky-Golay filters; together with two commonly use image filters: the Mean and Median filters, respectively.

Linear Prediction Coefficients approach. A suitable approach to remove the correlation among the samples of magnitude of MCLT of watermarked signal, is to use the autoregressive modeling scheme as shown in figure 4. Here the LPC removes the correlation of the watermarked information and deliveries a noise-like signal (Atal *et al*, 1971).

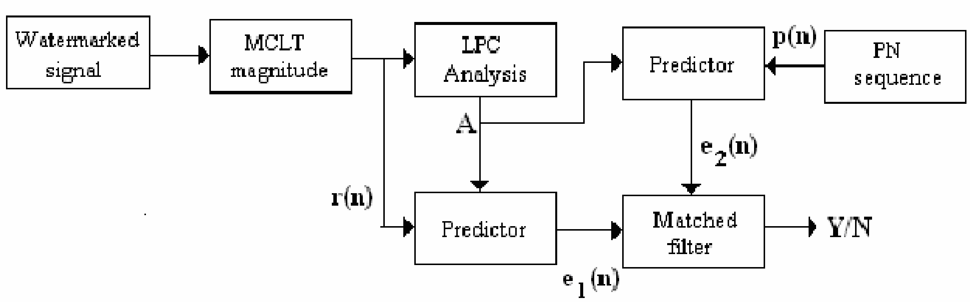


Fig. 4. Watermark LPC detecting scheme

Consider the watermarked audio signal which is generated by using the embedding scheme shown in figure 1. Thus to carried out the watermark detection, firstly the magnitude of MCLT of watermarked signal, *r(n)*, is estimated as shown in figure 4, which is given by the addition of the MCLT of original audio signal, *x(n)*, plus the watermark, *p(n)*. Next the magnitude of MCLT transformed watermarked signal is filtered by a linear predictor **A** filter. Then the output error signal $e_1(n)$ is given by

$$e_1(n) = \mathbf{A}^T \left(\mathbf{X}(n) + \alpha \mathbf{P}(n)\right) \tag{8}$$

where **X**(n) and **P**(n) are the input vectors of FIR linear predictors denoted by **A**. In a similar form the output error $e_2(n)$ is given by

$$e_2(n) = \mathbf{A}^T \mathbf{P}(n) \tag{9}$$

Assuming that **X**(n) and **P**(n) are uncorrelated among them, the cross correlation between $e_1(n)$ and $e_2(n)$ is given by

$$\phi_{e_1 e_2}(n) = E\left[e_1(n) e_2(n+k)\right] = E\left[p(n) p(n+k)\right] \tag{10}$$

Thus the system shown in figure 3 is able to detect, in a blind form the watermark embedded by using the watermarking system shown in figure 1.

Cepstrum approach. The Cepstrum procedure, proposed by Kirovski et al (2003), is other suitable approach for blind watermarking detection, which can be sumarized as follows: Firstly, the cepstrum of the magnitude of MCLT, **X**, is computed. Next, compute the discrete cosine transform of **X,** i.e.**, Y**=DCT(**X**). Next, vector **X** is high pass filtered using eq. (11), where k is the normalized cutoff frequency, /N and q is a constant.

$$\mathbf{Y}_c(n) = \begin{cases} 0 & n = 0,1,2,\ldots k \\ q\mathbf{Y}(n) & n = k+1 \ldots N \end{cases} \tag{11}$$

Subsequently the inverse DCT of the filtered signal if computed, $\mathbf{X}_c$=IDCT($\mathbf{Y}_c$); and finally, to carried out the watermark detection, the cross correlation between the sequence $\boldsymbol{X}_c$ and the watermark is computed. Here if the maximum of the cross correlation is larger than a given threshold the watermark is considered to be present.

Savitzky-Golay filter. The Savitzky-Golay filter is considered as one of the more efficient smoothing methods (Cvejic *et al*, 2001). This filter, which is derived directly from a particular formulation of the data-smoothing problem in time domain, is optimal in the sense that they minimize the least square error in fitting a polynomial to the noisy data frames. The watermarking detection process using the Savitzky-Golay filter is given as follows: a) compute the output signal of Savitzky-Golay filter which, for a five point smooth of data is given by

$$\mathbf{X}(i) = \left(-3\mathbf{Y}(i-2) + 12\mathbf{Y}(i-1) + 17\mathbf{Y}(i) + 12\mathbf{Y}(i+1) - 3\mathbf{Y}(i+2)\right)/35 \tag{12}$$

where ***Y(i)*** is the magnitude of *i-th* components of MCLT of watermarked signal and ***X(I)*** is the *I-th* component of MCLT smoothed vector. Next compute *e(i)* given by the difference between ***Y(i)*** and *X(i)*, i.e.

$$e(i) = \mathbf{Y}(i) - \mathbf{X}(i) \tag{13}$$

Finally the watermark detection is carried out by computing the cross correlation between *e(i)* and the watermark. This is then compared with a given threshold, *Th*, and, if the maximum of the cross correlation is larger than *Th* the watermark is considered to be present

Median Filter. The median filter, which has been widely used in image enhancement applications, is another attractive approach for whitening the watermarked MLCT of watermarked audio signal, because it is less expensive than the previously described methods. The median filter based MCLT whitening algorithm can be summarized as follows: Firstly, apply the median filter to the magnitude of MCLT of watermarked signal; where the median filter output is given by $X_S((N+1)/2)$ if N is odd; and $0.5(\mathbf{X}_S(N/2)+\mathbf{X}_S(N/2+1)$ if N is even, where $\mathbf{X}_s$ is a vector that consists of the MCLT coefficients of input signal in a decreasing order according their magnitude and N is the block size. The Median filter is a non linear filter which requires a window long enough to perform the whitening process. However, because it can not be theoretically estimated, after several experiments we found that with a window of 256 samples the Mean filter performs a fairly good whitening process. After the whitening process is performed, the difference, *e(n),* between the *n-th* filtered and original watermarked MCLT signal samples is computed. Finally, the correlation between *e(n)* and the watermark is computed. Here if the maximum of the cross correlation is larger than a given threshold, *Th*, the watermark is considered to be present.

Mean filter. The mean filter has also been widely used for noisy data smoothing because it has a lowpass frequency response. Then it can be used in a similar form that the Savitzky-Golay filter to whitening the MCLT magnitude of watermarked coefficients in order to keep a low computational complexity. Thus the watermarking detection using a mean filter based MCLT magnitude whitening algorithm is carried out as follows: Firstly compute the average of MCLT magnitude using a sliding window of length equal to 11 as follows

$$\mathbf{X}(n)=\frac{1}{11}\sum_{k=0}^{10}\mathbf{Y}(n+k), \quad n=1,2,....M \tag{14}$$

where M is the number of elements of the magnitude of MCLT of watermarked audio signal. Next the difference, *e(n)*, between the filtered sequence, ***X****(n),* and the MCLT watermarked one, ***Y****(n)* is computed. Next the cross correlation between *e(n)* and the known watermark is computed. Finally, if the maximum of the estimated cross correlation is larger than a given threshold, *Th*, the watermark is considered to be present.

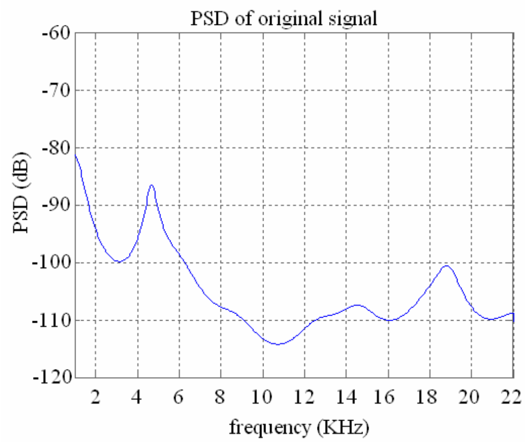
4 Real-Time Implementation

The proposed audio watermarking system is implemented in two state-of-the-art fixed point DSP (Digital Signal Processor) boards. These DSP boards, which has an operation frequency of 150 MHz, performs 1200 mega instructions per second (MIPS) and 600 mega floating point operations per second (MFLOPS). It has an audio codec, 16 MB SDRAM, 512 KB Flash memory and a host communication module. Thus to achieve a real time operation the computational complexity of proposed algorithm must be less than 1200 MIPS.

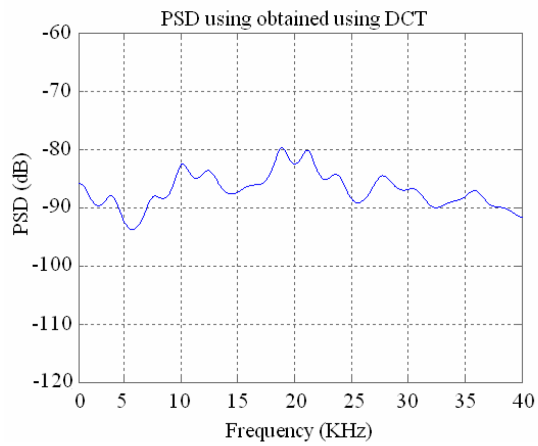
To analyze the computational complexity of proposed algorithm, consider the number MIPS required for implementing the proposed scheme. Thus the proposed algorithm requires estimate the MCLT coefficient which are given by $\mathbf{X}=\mathbf{P}_a^T\mathbf{x}$, where $\mathbf{P}_a$ is given by eqs. (1)-(3), (7) and the IMCLT given by $\mathbf{y}=\mathbf{P_S}\mathbf{X}$, where $\mathbf{P}_s$ is given by (4)-(6), together with watermark generation. To reduce the computational complexity and the MCLT and IMCLT can be computed using the fast algorithm for MCLT proposed by Malvar (2005) based on an optimized FFT implementation. Assuming a sampling rate of 44.1 KHZ and a block size of 2048 samples and optimized FFT requires approximately $21.13\text{x}10^4$ cycles, an IFT requires about $21.14\text{x}10^4$ cycles and the LFSR requires $55.53\text{x}10^5$ cycles, as well as several other operations. Thus the proposed algorithm requires, for watermark embedding, approximately $112\text{x}10^6$ cycles. Finally taking in account that the DSP TMS320C6416T performs 8 instructions per cycle, the computational complexity of proposed algorithm is about 896 MIPS which is fewer the 1200 MIPS that it is possible to perform.

Next consider the computational of the watermark detection stage; which requires the estimation of MCLT coefficients based on the FFT and a LFSR, described above, together with the whitening process which requires about $21.65\text{x}10^6$ cycles; and the correlation estimation. Thus the computational complexity of proposed watermark detection algorithm requires about 45.86x10^6 cycles. Thus taking in account that these DSP performs 8 instructions per cycle, the proposed watermark detection algorithm requires about 366.88 MIPS.

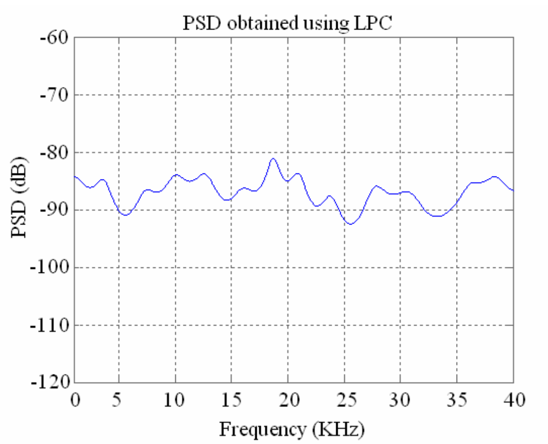
Finally the embedded and detection algorithms were also written in C-language to allow portability between different platforms, doing the system is host-independent by including an assembly routine which was written to boot the board with watermarking embedding/detection programs.



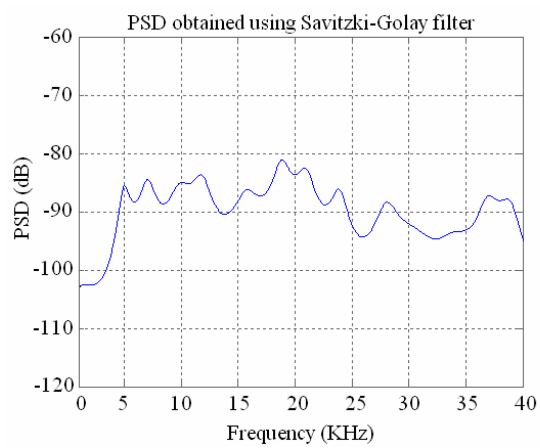
(a)



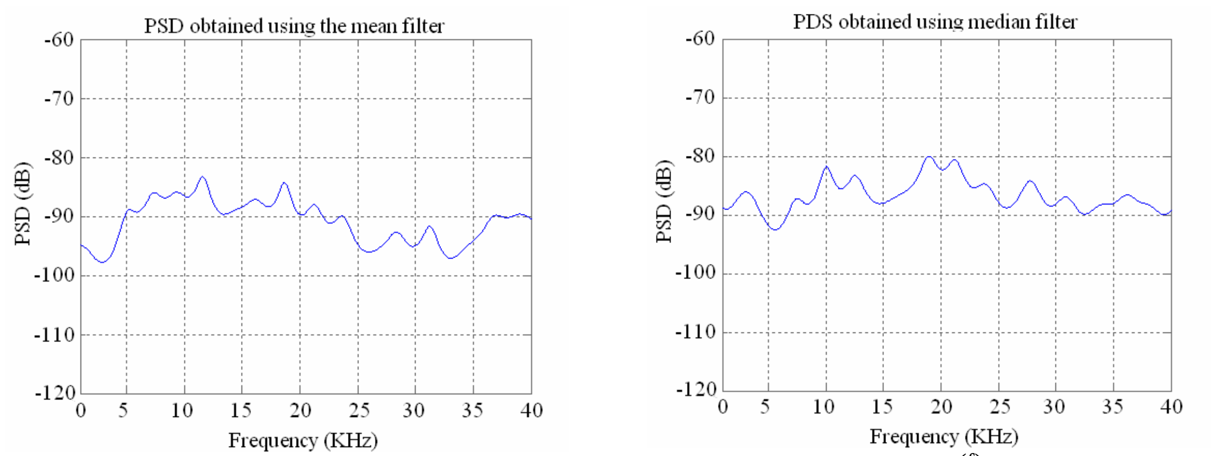
(c)



(b)



(d)



(e) (f)

Fig. 5 MCLT whitened signal using five different methods (a) original MCLT PSD, (b) PSD of whitened MCLT signal using the LPC approach, (c) PSD of whitened MCLT signal using the DCT, (d) PSD of whitened MCLT signal using the Savitzky-Golay filter, (e) PSD of whitened MCLT signal using the mean filter, (f) PSD of whitened MCLT signal using the median filter

5 System Evaluation

The performance evaluation of proposed audio watermarking system was carried out using the DSP implementation in which the audio signal is feed into a DSP for watermark embedding, then the watermarked audio signal is feed into loudspeaker as shown in Fig 5. Subsequently the watermarked audio signal is detected by a microphone and processed by a second DSP for watermark detection. Here the proposed watermark embedding implementation system was evaluated in terms of a) audio quality, b) detection performance and c) robustness to common attacks. Next sections present a detailed evaluation of proposed system using the above mentioned tests.

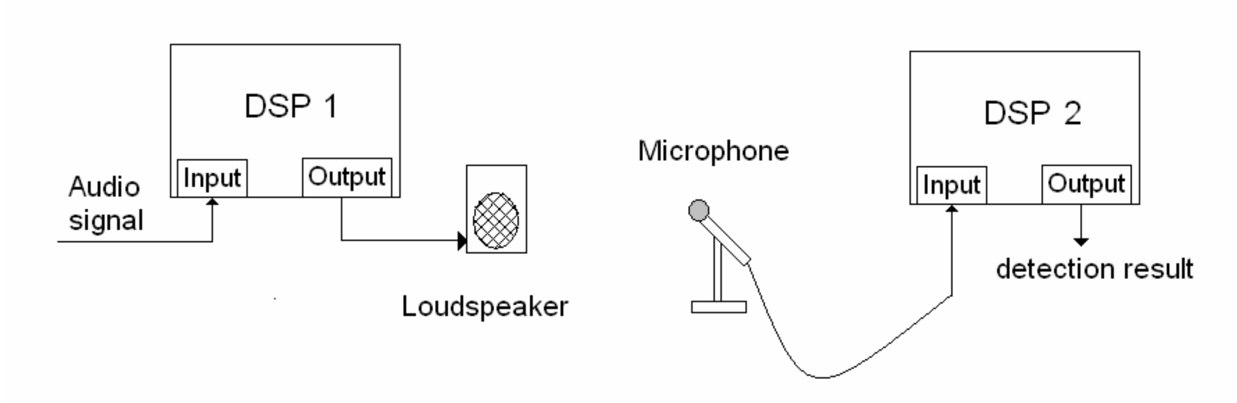


Fig. 6. Evaluation setup using DSP boards for real time implementation of proposed watermarking system

5.1 Audio quality

To evaluate the signal quality after the embedding watermarking process, two different tests were carried out: the mean opinion scoring and the modified Bark spectral distortion which are described next..

Mean Opinion Score (MOS). The MOS (Mean Opinion Scoring) test, regarding the watermark imperceptibility was carried out, where the criterions used are shown in Table 1. The MOS evaluation, whose results are shown in Table

2, was carried out with a population of 25 persons using 5 different kinds of music: Classic music, Rock music, Pop music, Instrumental music and Latin music.

Table 1. MOS evaluation criterion

Score	Watermark imperceptibility
5	Imperceptible
4	Perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Table 2. MOS evaluation results for table 1 criterion

Music kind	Score
Classic music	4.4
Rock music	4.6
Pop music	4.7
Instrumental music	4.3
Latin music	4.8

The MOS evaluation shows that embedding process is clear to the HAS and it is difficult to discriminate between the original signal and the watermarked one. However the MOS is a subjective criterion which depends on the population performing the evaluation, thus to obtain an objective evaluation the Modified Bark Spectral Distortion test was used which is described next.

Modified Bark Spectral Distortion (MBSD). The second test applied to watermarked signal is the Modified Bark Spectral Distortion, MBSD, (Yang et al, 1997 and 1998) which is given by

$$MBSD = \frac{1}{N}\sum_{j=1}^{N}\left[\sum_{i=1}^{K}\left|L_x^{(j)}(i) - L_y^{(j)}(i)\right|^2\right] \tag{15}$$

where $L_x^j(i)$ and $L_y^j(i)$ are the loudness of original and watermarked signals, estimated using the Bark spectrum, respectively; N is the number of frames and K is the number of critical bands. The MBSD estimates audio signal distortion in loudness domain taking in account the noise-masking threshold of the HAS in order to include only audible distortions in the estimation of the distortion measured and from, eq. (15) it must be as close to zero as possible. Here the MSBD test was carried out using 5 different kinds of music: Classic music, Rock music, Pop music, Instrumental music and Latin music whose results are shown in Table 3.

Table 3. MBSD evaluation results for five different kinds of music

Music kind	MBSD (dB)
Classic music	-62.3
Rock music	-65.3
Pop music	-62.8
Instrumental music	-61.7
Latin music	-62.5

From table 3 it follows that the distortion introduced by the proposed watermarking embedding system is low enough because in all cases the MBSD is below -60 dB

5.2 Detection performance

The watermark detection is carried out using the first point the cross-correlation between the watermarked audio signal, *y(i)*, and the watermark, *w(i)*, which can be estimated as follows

$$S = \frac{1}{N}\sum_{i=1}^{N} y(i)w(i) \tag{16}$$

where *y(i)* the whitened MCLT coefficients of audio clip and *w(i)* is the watermark which consists of a pseudorandom number with $E[w]=E[w^3]=0$ and $E[w^2]=E[w^4]=1$. Assuming that the watermark is uncorrelated with the audio signal, and that $y(i)=x(i)+\alpha w(i)$, where α is the watermark strength it follows that: $E[S]=0$ if watermark is not present and $E[S]=\alpha$ if watermark is present; while variances given by

$$\sigma_S^2 = \left(\frac{\mu_X{}^2 + \sigma_X{}^2 + \alpha^2}{N}\right)$$

(17)

if no watermark is present, and

$$\sigma_S^2 = \left(\frac{\mu_X{}^2 + \sigma_X{}^2}{N}\right) \tag{18}$$

if watermark is present, respectively, where μ_x is the mean of magnitude of MCLT of audio signal and σ_x^2 the variance of the magnitude of MCLT of audio signal. It can be seen that if the watermark length, N, increases, σ_S^2 decreases and then the estimator robustness grows. As mentioned in section 2.2, N=2048 is an adequate length of PN-sequence.

To evaluate the performance of proposed watermark detection algorithm, firstly the probability density function (pdf) of S was estimated, using two different hypotheses. The first hypothesis, H_1, is the probability that a watermark exists and the second one, H_2, is the probability that the watermark does not exist. To obtain these two pdf, 5000 watermark detections were carried out, using each one of the five whitening procedures presented above. Using the estimated PDFs, the detection threshold, Th, is set such that the probability of a false detection, Pfa, which is the area of H_2 in which the value of S is larger than Th, becomes smaller than 10^{-6}; while the probability of a misdetection, Pmd, which is the area of $H_{1,}$ in which the value of S is smaller than Th, is calculated from the empirical pdf. Figures 7-11 show the empirical histograms and threshold resulting of each whitening procedures tests, the histogram of the correlation between the watermark and a no-watermarked signal is denote by H_2 and localized on the left part of Th in each the figure, while the histogram of the correlation between the watermark and a watermarked signal, denoted by H_1 is located on the right side of Th in each figure. Here as mentioned before, Th was calculated for the specified Pfa. Finally the Table 4 shows the Pmd and threshold obtained during the detection tests of whitening processes.

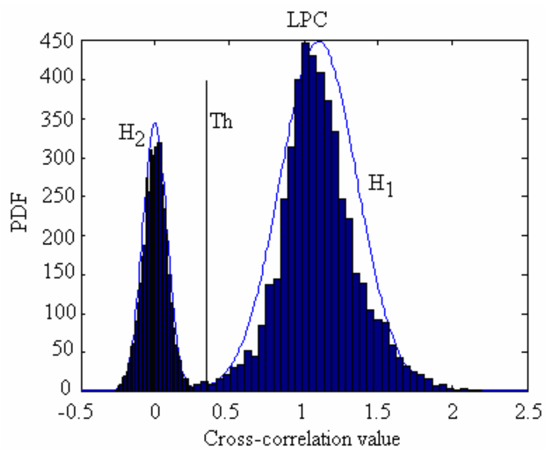


Fig. 7. Evaluation results of watermark detection using LPC detecting scheme

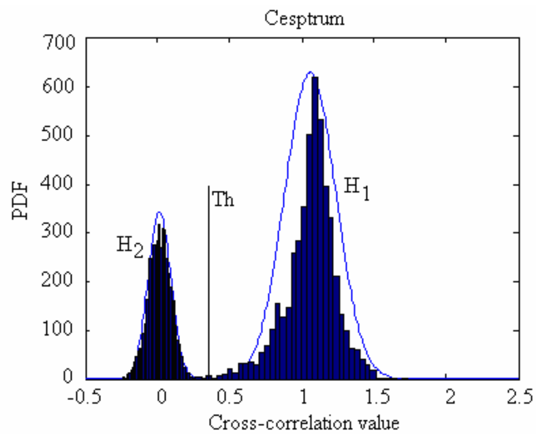


Fig. 8. Results of watermark detection using Cepstrum filter detecting scheme

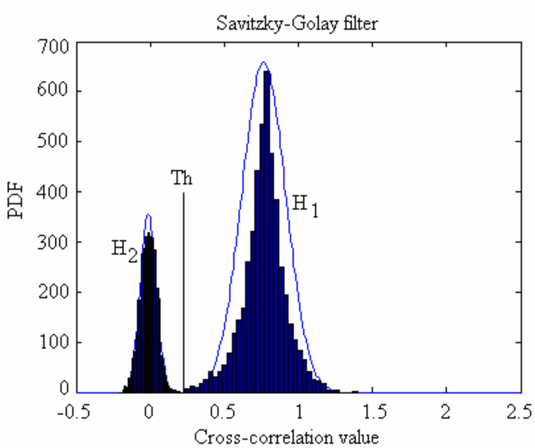


Fig. 9. Results of watermark detection using Savitzky-Golay filter detecting scheme

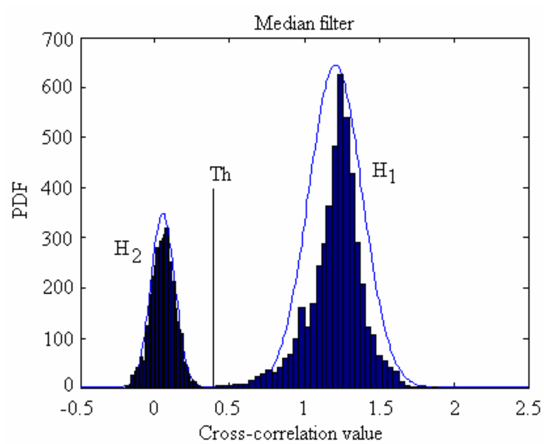


Fig. 10. Results of watermark detection using median filter detecting scheme

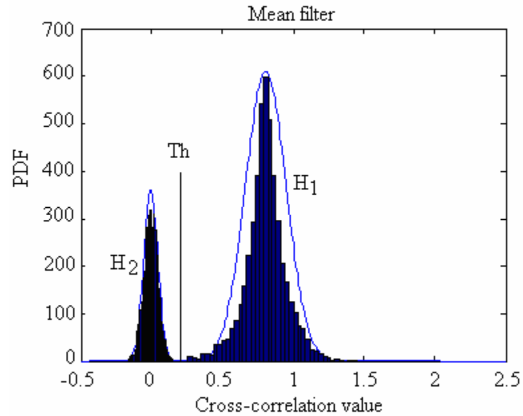


Fig. 11. Evaluation results of watermark detection using Mean filter detecting scheme

Table 4. Results of whitening procedures evaluations

Whitening method	LPC	Cepstrum	Savitzky-Golay filter	Median filter	Mean filter
Threshold	0.3464	0.35309	0.22124	0.39442	0.20626
Pmd	1.7231e-3	3.5243e-5	1.1772e-4	2.7686e-6	1.4493e-5

Table 5. Average of correlation test after attacks

Attacks	MP3	Noise addition	Echo addition	Band-pass filtering	All-pass filtering	Ogg coding	AAC coding
Correlation test	0.7739	0.8494	0.9111	0.7202	0.7796	0.8415	0.8602

The evaluation results show that the overlapping areas between H_1 and H_2 are negligible specially when the Mean and Median filter are used in comparison with those obtained using the LPC, Cepstrum and Savitzky-Golay, and as a consequence the probability of a correct detection is larger with the Mean and Median filter in comparison with the others three methods.

5.3 Robustness test

A set of typical attacks was applied to the proposed watermarking scheme (Garcia et al, 2005): firstly the watermarked audio signal is compressed using a MPEG coding with 64 kbps. Next the watermarked was corrupted with additive noise whose power as 32dB below the watermarked signal power. Next the watermarked signal was distorted with an echo signal whose delay was 125ms and its amplitude 40% with respect to the original watermarked signal. To test the robustness regarding filtering operations the watermarked signal was pass through a band-pass filtering with cutoff frequencies of 100Hz and 3500Hz, as well as through a nonlinear phase all-pass filtering. Finally a Ogg conversion with quality equal to 60% and, an AAC coding at 64 kbps were used for testing purposes. In all cases the watermark detection was carried out using the Median filter based whitening procedure. The table 5 shows the average of the correlation values between the watermark and watermarked signal obtained after these attacks. From this table it follows that in all cases the cross correlation value is much higher than the given threshold value (0.39442).

6 Conclusions

A MCLT based watermarking system using a blind detection approach is proposed. Because the blind detection scheme is assumed to be an AWGN, an efficient whitening method is required. To this end the proposed system was evaluated using five different whitening methods. These evaluation results show that, among the evaluated methods, the Median and median filters provide better whitening performance than other used whitening methods such as LPC, DCT and Savitzky-Golay filters. Evaluation about audio signal distortion was also performed using subjective and objective criterions such as MOS and MDSB, respectively. These evaluation results show that the distortion in the watermarked signal is negligible from the perceptual point of view, and then the watermarked signal is quite closed to the original one. Finally robustness was also evaluated using different kinds of attacks. These results show that the proposed system also provides robustness to several kinds of attacks, such as MP3 compression, additive noise, echo adding, filtering as well as Ogg and ACC coding. All evaluations were carried out using a real time implementation of proposed algorithm which requires 896 MPIS for watermark embedding and 366.88 MIPS for watermark detection which are smaller than the 1200MIPS that is able to perform the DSP TMS320C6416T.

References

1. **Atal B. and S. Hanauer**, “Speech analysis and synthesis by linear prediction of speech wave”,. *J. Acoustics Soc. Am.*, pp 637-655, 1971
2. **Bassia, P. and I. Pitas**. "Robust audio watermarking in the time domain", *IEEE Transactions on Multimedia, Vol.3, No 2.*, pp. 232-241, June 2001.
3. **Bender, W., D. Gruhl, N. Morimoto, A. Lu**. "Techniques for data hiding", *IBM System Journal, Vol. 35*, pp. 313-336, 1996.
4. **Cvejic, N. and T. Seppanen**, “Audio Prewhitening based on Polynomial Filtering for Optimal Watermark Detection” *Information processing laboratory*, University of Oulu, Finland, 2001.
5. **Cvejic, N. and T. Seppanen**, “Improving audio watermarking scheme using psychoacoustic watermark filtering”, *Information processing laboratory*, University of Oulu, Finland, 2001.
6. **Feller W.**, “An introduction to probability theory and its applications”, John Wiley & Sons, 1968.
7. **Garcia, J., M. Nakano, H. Perez**, “Real Time Implementation of Low Complexity Audio Watermarking Algorithm””, *Proc. Third International Workshop on Random Fields and Processing in Inhomogeneous Media*, pp. 7-9, Guanajuato, Mexico, October 2005.
8. **Haitsma, J., Michiel van der Veen, Ton Kalker, Fons Bruekers**. "Audio watermarking for monitoring and copy protection", *ACM Multimedia Workshop Marina Del Ray, USA*, pp.119-122, 2000.
9. **Kirovski, D. and H. Malvar**. “Robust covert communication over a public audio channel using spread spectrum”, *4th International Information Hiding Workshop*, Pittsburgh, USA, April 2001.
10. **Kirovski, D. and H. Malvar**. “Spread spectrum watermarking of audio signals”, *IEEE Transactions on Signal Processing, Vol.51, No.4*, pp. 1020-1033, April 2003.
11. **Malvar H**. “A modulated complex lapped transform and its applications to audio processing”, *ICASSP 99*, Phoenix, USA, March 1999.
12. **Malvar H**. “Fast algorithm for the modulated complex lapped transform”, *Technical Report*, Microsoft Research, January 2005 .
13. **Neubauer, C., J. Herre, K. Brandenburg**. "Continuous steganographic data transmission using uncompressed audio", *Information Hiding Second International Workshop*, Portland, USA, April 1998.
14. **Swanson, M., Bin Zhu, Ahmed H. Tewfik**. "Robust audio watermarking using perceptual masking", *Signal Processing, Vol. 66*, pp. 337-355, 1998.
15. **Yang, W., M. Benbouchta and R. Yantorno**, “Performance of the modified bark spectral distortion as an objective speech quality measure”, *ICASSP*, Vol. 1, pp 541-544, Seattle, 1998.

16. **Yang, W., M. Dixon and R. Yantorno**, "A modified bark spectral distortion measure which uses noise masking threshold", *IEEE Speech Coding Workshop*, pp.55-56, Pocono Manor, 1997.

Jose Juan Garcia Hernandez *received his Bachelor degree in December 2002, and his M. E. degree in December 2004, from the Mechanical and Electrical Engineering School of the National Polytechnic Institute of Mexico, where he is now Ph. D degree student. His research interests are in audio and speech signal processing, especially audio watermarking system.*

Mariko Nakano Miyatake *received a M.E. degree in Electrical Engineering from the University of Electro-Communications, Tokyo Japan in 1985, and her Ph. D in Electrical Engineering from The Universidad Autónoma Metropolitana (UAM), Mexico City, in 1998. From July 1992 to February 1997 she was with the Department of Electrical Engineering of the UAM Mexico. In February 1997, she joined the Graduate Department of The Mechanical and Electrical Engineering, School of The National Polytechnic Institute of Mexico, where she is now a Professor. Her research interests are in signal processing, watermarking, neural networks, pattern recognition and related field.*

Hector Pérez Meana *received his Ph. D. degree in Electrical Engineering from Tokyo Institute of Technology, Japan, in 1989. In February 1997, he joined the Graduate Department of The Mechanical and Electrical Engineering School of the National Polytechnic Institute of Mexico, where he is now a Professor. In 1991 he received the IEICE excellent Paper Award, and in 2000 the IPN Research Award. He is a senior member of the IEEE, member of the IEICE, The SNI level II and The Mexican Academy of Science. His principal research interests are adaptive filter systems, image processing, pattern recognition and related fields.*