Blind Tamper Detection in Audio using Chirp based Robust Watermarking

O. FAROOQ, S. DATTA, AND J. BLACKLEDGE
Department of Electronic and Electrical Engineering
Loughborough University, Loughborough
Leicestershire, LE11 3TU UK

o.farooq@lboro.ac.uk, http://www.lboro.ac.uk/departments/el/staff/farooq.html
s.datta@lboro.ac.uk, http://www.lboro.ac.uk/departments/el/staff/datta.html
jon.blackledge@btconnect.com, http://www.lboro.ac.uk/departments/el/staff/blackledge.html

Abstract In this paper, we propose the use of ‘chirp coding’ for embedding a watermark in audio data without generating any perceptual degradation of audio quality. A binary sequence (the watermark) is derived using energy based features from the audio signal and chirp coding used to embed the watermark in audio data. The chirp coding technique is such that the same watermark can be derived from the original audio signal as well as recovered from the watermarked signal. This not only enables the ‘blind’ recovery of the watermark, but also provides a solution for deriving two independent extraction processes for the watermark from which it is possible to ensure the authenticity of audio data and any mismatch indicating that the data may have been tampered with. To evaluate the robustness of the proposed scheme, different attacks such as compression, filtering, sampling rate alteration, for example, have been simulated. The results obtained reflect the high robustness of the watermark method used and is effectiveness in detecting any data tampering that may have occurred. For perceptual transparency of the watermark, Perceptual Assessment of Audio Quality (PEAQ ITU-R BS.1387) on Speech Quality Assessment Material (SQAM) has been undertaken and an average of -0.5085 Objective Difference Grade achieved.

Key-words— Chirp Coding, Robust Audio Watermarking, Self-Authentication, Tamper Detection, Wavelet Transform.

1. Introduction

The growth in digital multimedia technologies has made users switch from using analog to digital data. The use of internet and wireless applications has made it easy and fast to transmit data and with the availability of powerful computing, digital copying and tampering of data can be performed relatively easily. From the viewpoint of media producers and content providers, the ability to copy digital data without loss of fidelity is undesirable because it may compromise financial integrity and also violate copyright [1]. Editing/tampering the content of data may give misleading/wrong information. Therefore, the protection and enforcement of intellectual property rights for digital media has become an important issue. Cryptography can been used to guarantee secure transfer of data from point-to-point. However, once the data has been decrypted by a valid recipient, it can be subsequently re-distributed in its original form without the knowledge of the owner.

One solution to ownership protection is digital watermarking based on techniques for embedding a low energy signal (watermark) in an original multimedia data or host (i.e. audio, image and video data) in such a way that the perceptual quality is not degraded. This requires that the watermark be robust and that is can not easily be removed from the data/media unless there is an appreciable degradation in the perceptual quality. The corruption of a watermark (an ‘attack’) may achieved by using techniques such as lossy compression, scaling, signal cropping, re-sampling, re-quantization, etc. of the media.

To authenticate originality of the data, another class of watermarking is carried out known as fragile watermarking. In contrast to robust watermarking, in this case, the watermark breaks as soon as any processing is applied to the watermarked signal. Each watermarking technique depends on the application to which it is applied. Robust watermarks are generally used for copyright and ownership verification. In comparison, fragile watermarks are useful for the purpose of authentication and integrity attestation. Developing imperceptible audio watermark is difficult compared to image watermark because of the high sensitivity of the human ear.

Audio watermarking techniques exploit the fact that the human auditory system is insensitive to small amplitude changes, either in the time domain [1, 2], frequency domain [3, 4, 5] or other transform domains [6, 7, 8, 9, 10, 11, 12]. Usually, embedding is performed in high amplitude portions of the signal, either in the time or frequency domains taking advantage of the 'masking property' of the human ear. Another commonly used technique for watermarking is based on hiding a low-amplitude spread-spectrum (SS) sequence, which can be detected via correlation techniques [13, 14, 15].

All the watermarking techniques discussed and referenced above fall either into the category of robust or fragile watermarking. In this paper, we propose a signal dependent chirp based and robust audio watermarking technique which helps to detect tampering of the watermarked signal. This technique provides watermarks that are both robust and fragile. Section 2 of this paper gives the basic introduction to watermarking and the mathematical background to embedding and extracting the watermark. The proposed chirp based watermarking technique is discussed in Section 3 along with the watermark sequence extraction process. The watermark sequence is generated by calculating the energy in sub-bands obtained by using wavelet decomposition of the audio signal. To extract the watermark sequence from the audio data, the same wavelet decomposition is again carried out to evaluate the sub-band energies of the signal. The normalized sub-band energies are used as features to generating the watermark sequence. The scheme is found to be robust to the attacks simulated and can thus detect tampering in the audio signal.

2. Watermarking: Background

Methods of watermarking digital data have applications in a wide range of areas. Digital watermarking of images has been researched for many years in order to achieve methods which provide both anti-counterfeiting and authentication facilities [16]. A principal equation that underpins this technology is based on the fundamental model for defining a signal in general and is given by [17]

$$s = \hat{P}f + n \qquad (1)$$

where f is the information content for the signal, $\hat{P}$ is some linear operator, n is the noise and s is the output signal. This equation is usually taken to describe a stationary process in which the operator $\hat{P}$ is invariant of time and the noise n is characterized by stationary statistics (i.e. the probability distribution function of n is invariant of time). In cryptology the operation $\hat{P}f$ is referred to as the processes of 'diffusion' and the process of adding noise is referred to as the process of 'confusion'. Here the principal 'art' is to develop methods in which the processes of diffusion and confusion are maximized, an important criterion being that the output s should be dominated by the noise n which, in turn, should be characterized by maximum entropy.

Digital watermarking can be considered to form part of the same field of study, namely, covert communications. Being able to recover f from s provides a way of reconstructing the information content of the signal. If we consider f as being information that constitutes a 'watermark', and n is a host signal (an audio signal, for example) in which the watermark is embedded, then our problem is to recover the watermark from the host signal. If, in addition, it is possible to determine that a copy of s has been made leading to some form of data degradation and/or corruption that can be conveyed through an appropriate analysis of f, then a scheme can be developed that provides a check on: (i) the authenticity of the data s; (ii) its fidelity, [1, 18].

Formally, the recovery of f from s is based on the inverse process

$$f = \hat{P}^{-1}(s - n) \qquad (2)$$

where $\hat{P}^{-1}$ is the inverse operator. Clearly, this requires the field n to be known *a priori*. If this field has been generated by a pseudo random number generator, for example, then the seed used to generate this field must be known *a priori* in order to recover the data f. In this case, the seed represents the private key required to recover f. Further, if the process of confusion is undertaken in which the signal-to-noise ratio is set to be very low (i.e. $\|n\| >> \|\hat{P}f\|$), then the watermark f can be hidden covertly in the data n provided the inverse process $\hat{P}^{-1}$ is well defined and computationally stable. In this case, it is clear that the host signal or image n must be known in order to recover the watermark f leading to a private watermarking scheme in which the field n represents a key.

Another approach is to consider the case in which the field n is unknown and to consider the problem or extracting the watermark f in the absence of this field. In this case, the reconstruction is based on the result

$$f = \hat{P}^{-1}s + m \tag{3}$$

where

$$m = -\hat{P}^{-1}n \tag{4}$$

Now, if a process $\hat{P}$ is available in which $\|\hat{P}^{-1}s\| >> \|m\|$, then an approximate (noisy) reconstruction of f can be obtained in which the noise m is determined by the original signal-to-noise ratio of the data s and hence, the level of covertness of the diffused watermark $\hat{P}f$. In this case, it may be possible to post-process the reconstruction (de-noising for example) and recover a relatively high-fidelity version of the watermark, i.e.

$$f \sim \hat{P}^{-1}s \tag{5}$$

This approach does not rely on a private key (assuming $\hat{P}$ is not key dependent). The ability to recover the watermark only requires knowledge of the operator $\hat{P}$ (and its inverse) and post-processing options as required. The problem here is to find an operator that is able to recover the watermark effectively in the presence of the field n. Ideally, we require an operator $\hat{P}$ with properties such that $\hat{P}^{-1}n \rightarrow 0$.

2.1 The Matched Filter

The matched filter (e.g. [19, 20, 21]) is a result of finding a solution to the following problem: Given that

$$s(t) = p(t) \otimes f(t) + n(t) \tag{6}$$

where

$$(p \otimes f)(t) \equiv \int p(t-\tau)f(\tau)d\tau, \tag{7}$$

find an estimate for the Impulse Response Function (IRF) *f* given by

$$\hat{f}(t) = q(t) \bullet s(t) \equiv \int q(t+\tau)s(\tau)d\tau \tag{8}$$

where

$$r = \frac{\left|\int Q(\omega)P(\omega)d\omega\right|^2}{\int |N(\omega)|^2 |Q(\omega)|^2 d\omega} \tag{9}$$

is a maximum and where the integrals are taken over the extent of each function. Here, Q, P and N are given by the Fourier tranforms of *q* (the time-domain filter), *p* (the instrument function) and *n* (the noise) respectively. The ratio defining *r* is a measure of the signal-to-noise (SNR) ratio. In this sense, the matched filter maximizes the signal-to- noise ratio of the output. Assuming that the noise $n(t)$ has a ‘white’ or uniform power spectrum, the (Fourier domain) filter *Q(ω)* which maximizes the SNR defined by *r* is given by

$$Q(\omega) = P^*(\omega) \tag{10}$$

and the solution is therefore

$$\hat{f}(t) = \frac{1}{2\pi}\int P^*(\omega)S(\omega)exp(i\omega t)d\omega \tag{11}$$

Using the correlation theorem it can be shown that

$$\hat{f}(t) = p(t) \bullet s(t) \tag{12}$$

The matched filter is therefore based on correlating the signal $s(t)$ with the instrument function $p(t)$. This filter is frequently used in systems that employ linear frequency modulated (FM) pulses - ‘chirped pulses’ which will be discussed later.

The value of *r* can be maximized when

$$Q(\omega) = \frac{P^*(\omega)}{|N(\omega)|^2} \tag{13}$$

If the noise $n(t)$ is ‘white’, then its power spectrum $|N(\omega)|^2$ is uniformly distributed. In particular under idealized conditions,

$$|N(\omega)|^2 = 1 \quad \forall \omega, \tag{14}$$

then

$$Q(\omega) = P^*(\omega) \tag{15}$$

2.2 Deconvolution of Frequency Modulated Signals

The matched filter is frequently used in systems that utilize linear frequency modulated (FM) pulses known as chirps or chirped pulses. Examples of where this particular type of pulse is used include real and synthetic aperture radar [22, 23], active sonar and some forms of seismic prospecting for example. Interestingly, some mammals (dolphins, whales and bats for example) use frequency modulation for communication and detection. The reason for this is the unique properties that chirps provide in terms of the quality of extracting information from signals with very low signal-to-noise ratios and the simplicity of the process that is required to do this (i.e. correlation). The invention and use of chirps for man made communications and imaging systems dates back to the early 1960s (the application of FM to radar for example); mother nature appears to have 'discovered' the idea some time ago.

The FM linear pulse is given (in complex form) by

$$p(t)=\exp\left(-i\alpha t^2\right), \qquad |t|\le T/2 \qquad (16)$$

where α is a constant and *T* is the length of the pulse. The phase of this pulse is αt^2 and the instantaneous frequency is given by:

$$\frac{d}{dt}\left(\alpha t^2\right)=2\alpha t \qquad (17)$$

which varies linearly with *t*. Hence, the frequency modulation is linear which is why the pulse is referred to as a linear FM pulse. In this case, the signal that is recorded is given by (neglecting additive noise)

$$s(t)=\exp\left(-i\alpha t^2\right)\otimes f(t) \qquad |t|\le T/2 \qquad (18)$$

In matched filtering we have

$$\begin{aligned}\hat{f}(t)&=\exp\left(i\alpha t^2\right)\bullet\exp\left(-i\alpha t^2\right)\otimes f(t)\\&=T\exp\left(i\alpha t^2\right)\sin c(\alpha Tt)\otimes f(t) \qquad |t|\le T/2\end{aligned} \qquad (19)$$

In some systems, the length of the linear FM pulse is relatively long. In such cases,

$$\begin{aligned}&\cos\left(\alpha t^2\right)\sin c(\alpha Tt)\cong\sin c(\alpha Tt) \quad and\\&\sin\left(\alpha t^2\right)\sin c(\alpha Tt)\cong 0\end{aligned} \qquad (20)$$

and so

$$\hat{f}(t)\cong T\sin c(\alpha Tt)\otimes f(t) \qquad (21)$$

Now, in Fourier space, this last equation can be written as

$$\hat{F}(\omega)=\begin{cases}\dfrac{\pi}{\alpha}F(\omega), & |\omega|\le\alpha T\\ 0, & otherwise\end{cases} \qquad (22)$$

The estimate $\hat{f}$ is therefore a band limited estimate of f whose bandwidth is determined by the product of the chirping parameter α with the length of the pulse *T*. Now, given that

$$s(t)=\exp\left(-i\alpha t^2\right)\otimes f(t)+n(t) \qquad (23)$$

after match filtering we obtain the estimate

$$\hat{f}(t)\cong T\sin c(\alpha Tt)\otimes f(t)+\exp\left(i\alpha t^2\right)\bullet n(t) \qquad (24)$$

The correlation function produced by correlating $\exp\left(i\alpha t^2\right)$ with $n(t)$ will in general be relatively low in amplitude since *n(t)* will not normally have features that match those of a chirp. It is therefore reasonable to assume that

$$\left\|T\sin c(\alpha Tt)\otimes f(t)\right\| >> \left\|\exp\left(i\alpha t^2\right)\bullet n(t)\right\| \qquad (25)$$

and that in practice, $\hat{f}$ is a band-limited reconstruction of *f* with high SNR. Thus, using chirps with matched filtering for the purpose of reconstructing an input in the presence of additive noise provides a relatively simple and computationally reliable method of 'diffusing' and reconstructing information encoded in the input function *f*. The ability for the matched filter to accurately recover information from linear FM type signals with very low SNRs leads naturally to consider its use for covert information embedding. This is the basis for the chirp coding method discussed in this paper - covertly watermarking digital signals for the purpose of signal authentication.

3 Chirp Based Watermarking

Chirp signals are used in pulse compression Real and Synthetic Aperture Radar applications to give low side-lobes when correlated with themselves. This gives the advantage of being detected in the presence of high background noise. In the watermarking application considered here, this background noise is the audio signal to be watermarked. The basic model for watermarking a signal $n(t)$ by a chirp *chirp(t)* is given by

$$s(t)=chirp(t)\otimes f(t)+n(t) \qquad (26)$$

where $s(t)$ is the watermarked signal and $f(t)$ the watermarking code. The instantaneous frequency of a logarithmic chirp signal is given by:

$$freq_i(t)= freq_0 + 10^{\beta t} \qquad (27)$$

where

$$\beta=log_{10}(freq_1 - freq_0)/ t_1 \qquad (28)$$

$freq_0$ is the initial frequency and $freq_1$ is the final frequency at time t_1. For the case of a logarithmic chirp, the final frequency should be greater than the initial frequency. Fig.1 shows three different types of chirps, classified in terms of the rate of change of frequency.

For authentication of the signal $s(t)$, two basic criteria must be satisfied: (i) $f(t)$ can be reconstructed accurately and robustly; (ii) it should be very sensitive to any degradation in the signal $s(t)$. The degradation in the signal may be due to lossy compression, filtering operations, re-sampling etc. To satisfy the

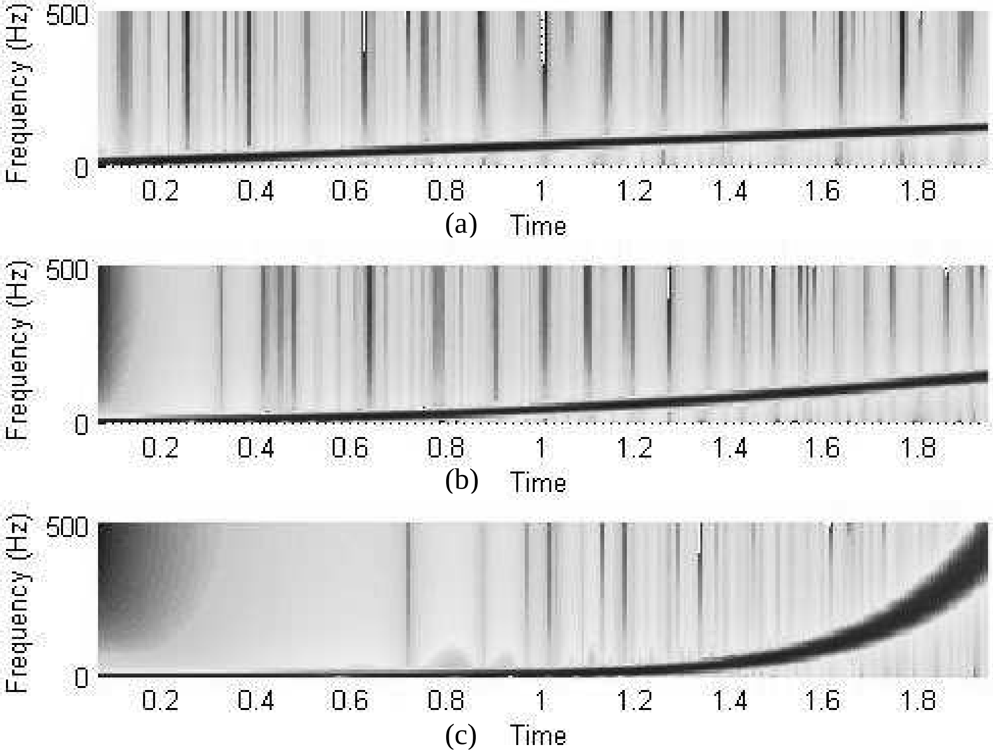

Fig. 1: Spectrogram showing the frequency sweep produced by a linear chirp (a), a quadratic chirp (b) and a logarithmic chirp (c).

first condition it is reasonable to consider $f(t)$ in terms of its representation of a bit stream, i.e. a digitized version of $f(t)$. This binary code can be generated by using a key or a set of keys, which, when reconstructed, is compared to the key(s) for the purpose of authentication of the data. However, this requires the distribution of the keys (public and/or private). Instead, a new technique for the generation of the binary sequence using the spectral characteristic of the signal itself is proposed. Once the binary sequence is generated, chirp coding can be applied. However, binary codes generated by using a key can also be used for chirp based watermarking which gives a robust watermark.

3.1 Watermark sequence generation

The entire power spectrum is decomposed into N sub-bands i.e.

$$P_i(\omega)=P(\omega), \qquad \omega\in[\Omega_{i-1},\Omega_i) \quad 1\le i\le N \qquad (29)$$

It is important that the signal $n(t)$ is band-limited and has a bandwidth of Ω_N. The set of functions $P_1,P_2,...,P_N$ represents the complete spectral characteristics of the signal $n(t)$. Since each of the components represents a unique part of the spectrum, a natural measure is to consider energy, which is determined by the integral of this function over the frequency range. The energy calculated in each sub-band is represented as a percentage of the total energy of the signal. The reason of calculating the percentage energies of the sub-bands is to avoid any influence of signal scaling on the authentication of the signal. The energy in the ith sub-band is given by:

$$E_i=\frac{100}{E}\int_{\Omega_{i-1}}^{\Omega_i} P_i(\omega)d\omega \qquad 1\le i\le N \qquad (30)$$

where

$$E=\int_0^{\Omega_N} P(\omega)d\omega \qquad (31)$$

An audio signal is split into sub-bands by applying the wavelet transformation which is defined by [24]

$$\hat{W}[f(t)]=F_L(t)=\int f(\tau)\varpi_L(t,\tau)d\tau \qquad (32)$$

where

$$\varpi_L(t,\tau)=\frac{1}{\sqrt{|L|}}\varpi\left(\frac{t-\tau}{L}\right) \qquad (33)$$

The wavelet transform is essentially a convolution transform in which $\varpi(t)$ is convolution kernel, but with

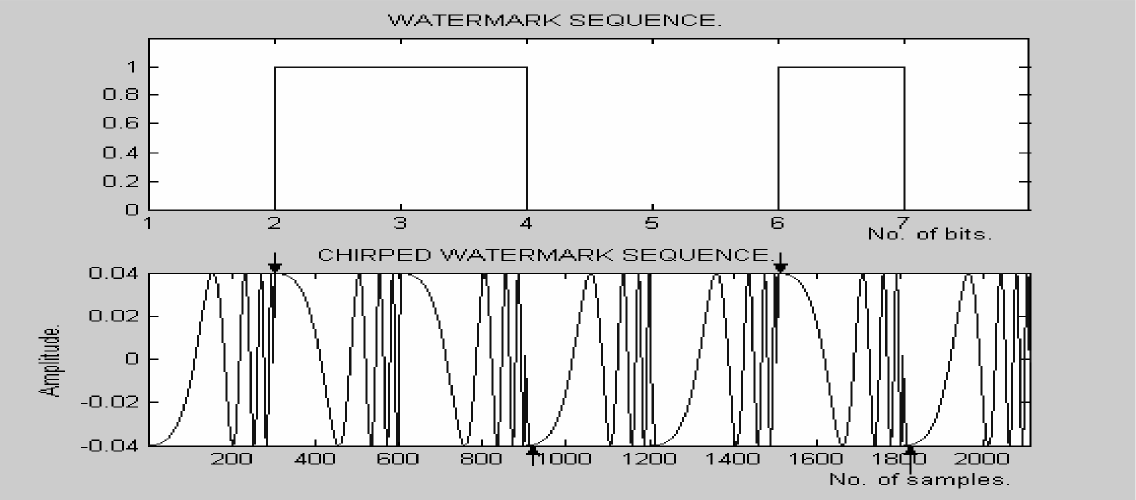

Fig. 2: Watermark sequence (0110010) and the corresponding chirp coded signal.

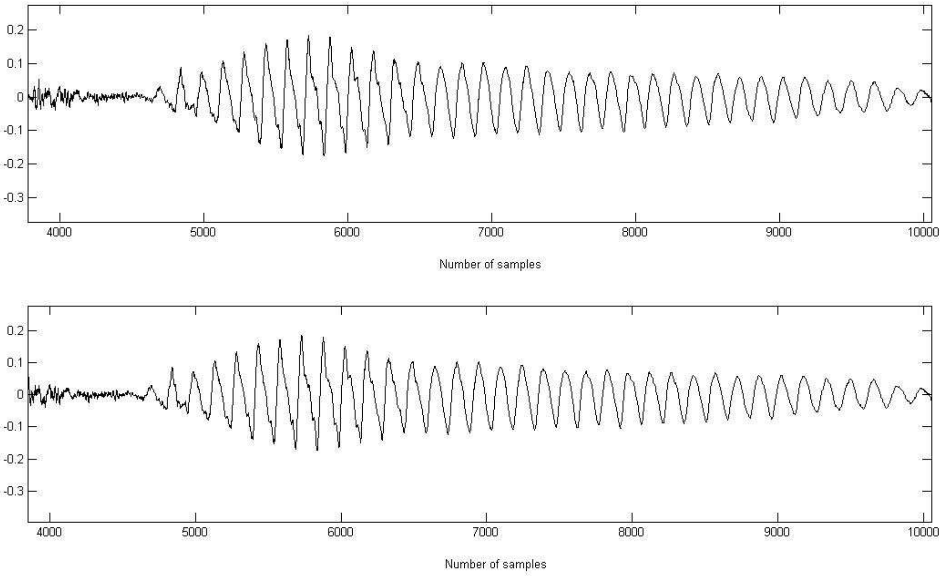

Fig. 3: A section of an original audio (above) and chirp based watermarked signal (below).

a factor L introduced. Introduction of this factor provides dilation and translation properties into the convolution integral (which is now a function of L).

This provides the transform with the ability to analyse signals at different resolutions.

The code generating method is based on computing the energies of the wavelet transformation over N levels. Thus, the signal $f(t)$ is decomposed into wavelet space to yield the following set of functions:

$$F_{L_1}(\tau), F_{L_2}(\tau), ...F_{L_N}(\tau) \tag{34}$$

The (percentage) energies of these functions are then computed, i.e.

$$E_i = \frac{100}{E} \int \left| F_{L_i}(\tau) \right|^2 d\tau \text{ where } E = \sum_{i=1}^{N} E_i \tag{35}$$

These energies represent the basic 'signature' of the audio signal from which the watermark is composed.

3.2 Watermark Embedding

Concatenating the total energy and sub-band energies provides the watermark feature or vector. The total energy, along with the percentage energy is converted into binary form using a 'b' bit representation. Concatenating all these bits into a single string gives the watermark sequence to be embedded in the audio signal. Fig. 2 shows a binary sequence (0110010) and the corresponding chirp code in which the phase of chirp is reversed for a '0'. The chirp frequency is taken to be low (upto 100 Hz) so that the embedded watermark is imperceptible with

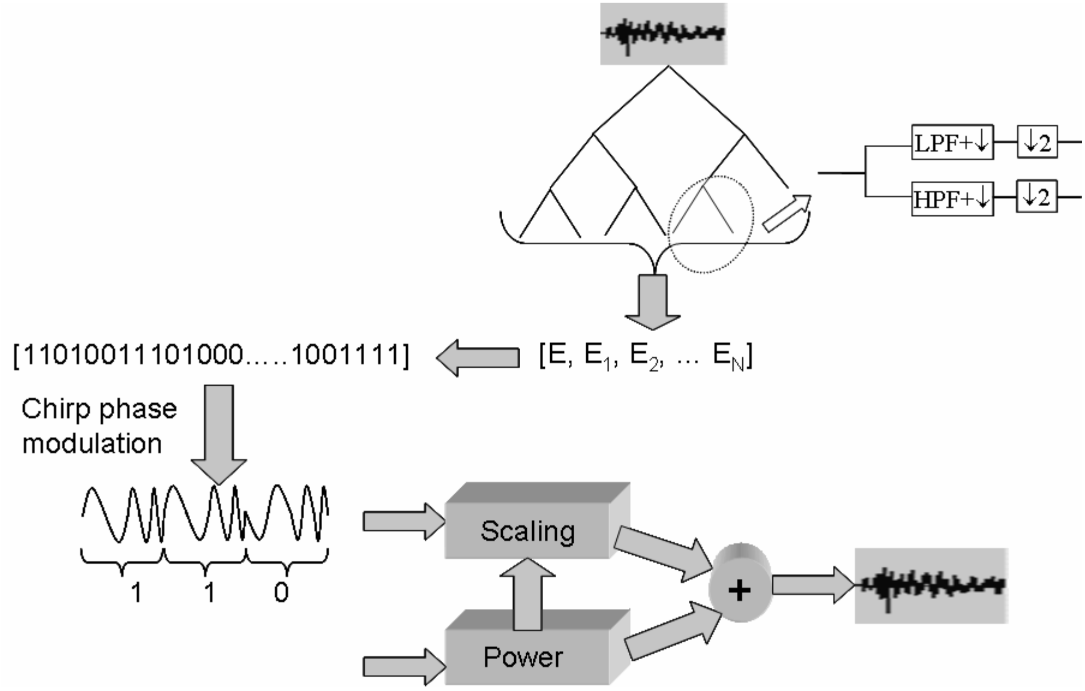

Fig. 4: Watermark sequence generation and chirp based watermark embedding.

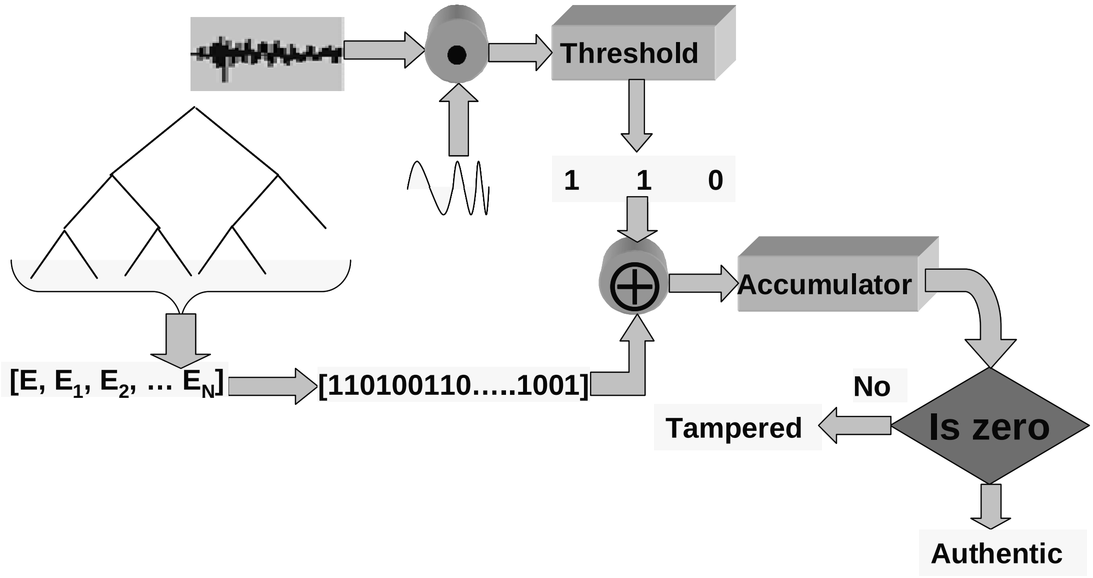

Fig. 5: Block diagram of watermark extraction by two independent process with tamper assessment.

regard to the low sensitivity of human ear in this frequency range. A scaled version of the chirp code is added to the audio signal to be watermarked. A section of original audio and the corresponding chirp based watermarked audio is shown in Fig. 3 and the watermarking scheme is shown in Fig. 4.

3.3 Watermark Recovery

The watermark is extracted using two processes. First, the watermarked audio signal is correlated with the chirp signal. The correlation obtained is then passed through a zero threshold. For positive values a ‘1’ is output and for negative values a ‘0’ is generated. This process recovers the watermark sequence. The second method used for the extraction of the watermark is based on the decomposition of the signal into sub-bands to calculate the percentage energies as explained in Section 3.1. The recovered watermark obtained by these two processes is compared using and EXOR operation. If the two sequences match a string of zeros is obtained and the audio is classified as being authenticated or else tampered. This watermark extraction scheme and check on authenticity is shown in Fig. 5.

4 Experimental Results

Different audio files from Speech Quality Assessment Material (SQAM) [25] were used for embedding the chirp based watermark. These audio signals are sampled at 44.1 kHz with 16 bits per sample. Logarithmic frequency sweeps were generated from 1 Hz to 100 Hz to make the chirp imperceptible. To calculate the watermark sequence, a wavelet transform is applied to the audio signal. The decomposition of the signal is carried out producing a binary tree structure as shown in Fig. 4. A total of seven sub-bands are achieved for calculation of energy. The watermark sequence was generated using percentage sub-band energies and total energy as explained in Section 3.1.

To further verify the above results, tests were undertaken based on Perceptual Assessment of Audio Quality (PEAQ, Basic) [26]. The average signal to watermark ratio achieved for the 16 audio signals used was in excess of 25 dB. Since the human ear sensitivity is more than 20 dB lower than the maximum sensitivity (which is around 3 kHz), the embedded chirp is not perceived at this signal to watermark ratio. The PEAQ algorithm is the ITU-R recommendation (ITU-R BS.1387) for perceptual evaluation of wide-band audio codecs. This algorithm models fundamental properties of the auditory system along with physiological and psychoacoustic effects. It uses both original and test signals, and applies techniques to find differences between them. An Objective Difference Grade (ODG) is evaluated using a total of eleven Model Output Variables (MOV) of the basic version of PEAQ. The original signal and watermarked signal for different embedding levels was used to evaluate ODG with results as given in Table 1. ODG values mimic the listening test ratings and have values -4.0 (very annoying) to 0 (imperceptible difference).

Table 1: Average values of the MOV of PEAQ basic model for audio taken from SQAM showing the imperceptibility of the proposed watermarking scheme

Model Output Variables	Scores
Total Noise to Mask Ratio	-14.9094
RMS Noise Loudness	0.383418
Relatively Disturbed Frames	0.014487
Objective Difference Grade	-0.50856

Although all the MOVs were calculated from the PEAQ (basic version) test, only those relevant to watermarking are reported here. The Noise-to-Mask Ratio (NMR) is an estimate in dBs of the ratio between the actual distortion (caused due to the embedding watermark in this case) and the maximum inaudible distortion. The total NMR is the average of the NMRs calculated over all frames. Negative NMR values indicate inaudibility whereas values larger than 0dB indicate audible distortions caused by the watermark.. This is an important test for checking the inaudibility of the embedded watermark at different levels. As stated earlier, the SNR obtained is in excess of 25 dB and shows a high level of watermark, but due to its very low frequency, it is not audible.

The noise loudness quantifies the partial loudness of distortions that is introduced when the watermark is embedded in the host signal. The Root Mean Square (RMS) value of noise loudness has a maximum limit of 14.8197. The average RMS noise loudness achieved by embedding the watermark is 0.383418 which indicates no perceptual distortion.

It is possible for the total NMR to be below 0dB (implying inaudibility), but there may be a large number of frames with small positive values and few frames with large negative values. This distribution can be seen by evaluating the number of disturbed frames. A relatively disturbed frame is one in which the maximum NMR exceeds 1.5 dB expressed as a fraction of the total frames. The results show that with chirp embedding, less than 1.5 percent of the frames have an NMR above 1.5dB.

Using the eleven different MOV of the basic PEAQ model, the average ODG for the audio files was found to be -0.50856 which is, in effect, equivalent to giving an imperceptible difference.

4.1 Robustness evaluation towards attacks

In order to evaluate the robustness of the proposed

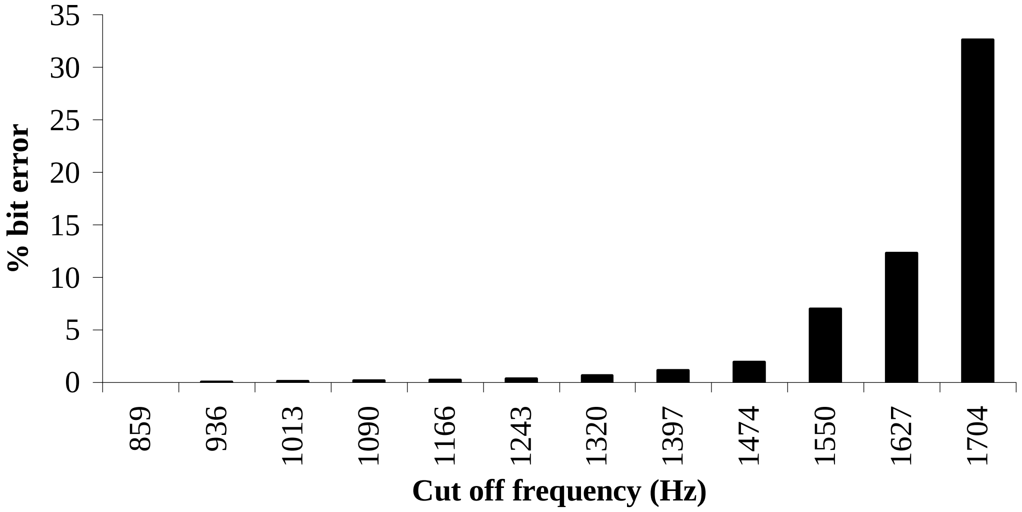

Fig. 6: Percentage bit error achieved for different cut-off frequencies of a high-pass filter.

scheme, different attacks were carried out on the watermarked signal. The watermark sequence was fully recovered in the presence of additive white Gaussian noise up to a level of 20 dB. Attacks constituting scaling, resampling, decimation and interpolation on the watermarked signal and did not result in errors in the recovering watermark sequence.

4.1.1 Filtering Attack

The embedded watermark can sustain limited high pass filtering attacks. A FIR filter of order 50 with different cut-off frequencies was designed to attack the watermarked audio signal. The percentage of bit s in error for different cut-off frequencies was evaluated and the results obtained are shown in Fig. 6. It is clear that for a cut-off frequency of upto 1KHz, the watermark can be fully recovered; however, for higher cut-off frequencies, errors start to occur. Since a filter has smooth transitions from stop-band to pass-band, the embedded chirp signal is not removed but is severely attenuated. Since the chirp can be extracted from a high noise background (as explained in Section 2), it is possible to extract the watermark at a high cut-off frequency. Further, it is important to note that by using filters with cut-off frequency higher than 1 kHz, the perceptual quality of the audio is degraded. Thus, the embedded watermark cannot be removed from the audio signal without appreciable degradation in its quality. However, there is no error in the recovered watermark subject to a low pass filtering attack due to the very low frequency content of the chirp.

4.1.2 MPEG layer 3 compression

A compression attack is one of the simplest and most common attacks associated with audio data. Since chirp based watermarking was carried out on a wave file format, it was necessary to evaluate the robustness towards compression attack. Experiments were conducted for a wide range of constant bit rate (CBR) ranging from 320 kbps to 128kbps. Further, variable bit rate (VBR) mp3 compression with lowest bit rates in the range of 32 kbps-128 kbps was also tested. Reconstruction of the watermark survived MPEG layer 3 compression with zero bit errors.

4.1.3 Signal Cropping

Arbitrary cropping samples of the watermarked signal was undertaken and the watermark detection process applied. Depending upon the length of the crop and the portion of the signal to which it was applied, different results were observed with regard to the evaluation of robustness. If data is removed from the end of the audio signal, then the watermark, prior to removal of the data, is fully detected. However, if the signal is cropped from the start or middle of the audio file, then the watermark is not detectable from the point beyond which the signal has been cropped. This is due to the fact that the offset parameter τ changes after cropping and hence the correlator cannot detect the watermark.

4.2 Tamper Detection

For tamper detection, the watermark sequence is

TABLE 2
TAMPERING ASSESSMENT UNDER DIFFERENT ATTACKS

Type of Attack	% bits in error	Tampered
Low-pass filtering (f_c=0.99)	31.25	Yes
High-pass filtering (f_c=0.01)	27.50	Yes
Up-sampling	53.57	Yes
Down-sampling	50.00	Yes
Re-sampling	17.50	Yes
Compression-Decompression	45.00	Yes

recovered by two methods from the watermarked audio signal. The first method is based on using the correlation scheme as explained in Section 3.3. The second method of extracting the watermark is by evaluating the total and sub-band energies obtained by wavelet decomposition (explained in Section 3.1). If the two sequences are the same, then the audio signal is self-authenticated, else it is tampered. The result of percentage bits in error due to different attacks on the watermarked audio is given in Table 2. Both low-pass (f_c being the normalized cut-off frequency) and high-pass filters were designed with 99% pass-band of the original signal. These filtering attacks resulted in negligible perceptual distortion but were detected easily because of changes in the sub-band energies distribution. For the same reason, up-sampling and down-sampling attacks were also detected. The wave file was compressed and then decompressed using mp3 CBR at 256 kbps which could be detected by the proposed scheme. The attack due to cropping could be detected since this attack reduces the energy in some of the sub-bands. This results in a mismatch between the two extracted watermarks thereby detecting tampering in the audio signal.

5 Conclusions

The ability for a chirp-coded watermarking to be detected in a high noise background is used here to embed a watermark in an audio signal. The use of a signal dependent watermark sequence helps in blind tamper detection of the audio signal. The scheme is found to be robust to attacks such as compression, scaling, resampling interpolation and decimation. While the scheme is robust to low pass filtering, it shows limited robustness to high pass filtering attacks. However, it is important to note that, the introduction of errors in the recovered watermark, leads to an audio quality that is substantially degraded due to the filtering. Since any attack on the watermarked audio will result in disturbing the original sub-band energy distribution, tampering can easily be detected for all the attacks carried out. Thus the proposed scheme has the dual advantage of being robust to various attacks as well as having a self-authentication capability to ascertain originality.

References

[1] F. A. P. Petitcolas, R. J. Anderson and M. G. Kuhn, Information Hiding – A survey, *Proc. of IEEE*, Vol. 87, No. 7, July 1999, pp. 1062-1078.

[2] A. N. Lemma, J. Aprea, W Oomen and L. Kerkhof, Temporal domain Audio Watermarking Technique, *IEEE Trans. on Signal Processing*, Vol. 51, No. 4, April 2003, pp. 1088-1097.

[3] J. Cox, J. Kilian T. Leighton and T. Shamoon, Secure spread spectrum watermarking for multimedia, *IEEE Trans. On Image Processing*, Vol. 6, No. 12, Dec. 1997, pp. 1673-1687.

[4] M. D. Swanson, B. Zhu, A.H. Tewfik, L. Boney, Robust audio watermarking using perceptual masking, *Signal Processing* 66 1998, pp. 337–355.

[5] D. Kirovski and H. S. Malvar, Spread Spectrum Watermarking for Audio signals, *IEEE Trans. on Signal Processing*, Vol. 51, No. 4, April 2003, pp. 1020-1033.

[6] D. Kundur and D. Hatzinakos, Digital Watermarking for Tell-Tale Tamper Proofing and Authentication, *Proceedings of the IEEE*, Vol. 87, 1999, pp. 1167-1180.

[7] Wu Shaoquan, Huang Jiwu, Huang Daren, and Q. Shi Yun, Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data

Transmission, *IEEE Trans. on Broadcasting*, Vol. 51, No. 1, March, 2005, pp. 69-76.

[8] M. Ketcham, and S. Vongpradhip, Intelligent Audio Watermarking using Genetic Algorithm in DWT Domain, *International Journal Of Intelligent Technology* Vol. 2, No. 2, 2007 pp. 135-140.

[9] Yi ju Wu and S. Shimamoto, A Study on DWT-Based Digital Audio Watermarking for Mobile Ad Hoc Network, *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, Vol. 2, 2006, pp. 247 – 251.

[10] R. Vieru, R. Tahboub, C. Constantinescu and V Lazarescu, New results using the audio watermarking based on wavelet transform, *International Symposium on Signals, Circuits and Systems*, Vol. 2, 2005, pp. 441-444.

[11] Xiaomei Quan and Hongbin Zhang, Audio watermarking based on psychoacoustic model and adaptive wavelet packets, *Proc. of 7th Int. Conf. on Signal Processing*, 2004 Vol. 3, pp. 2518-2521.

[12] Xiang-Yang Wang and Hong Zhao, A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT, *IEEE Transactions on Signal Processing*, Vol. 54, Issue 12, 2006, pp. 4835-4840.

[13] D. Kirovski and H. S. Malvar, Spread-spectrum watermarking of audio signals, *IEEE Trans. on Signal Processing*, Vol. 51, Issue 4, April 2003, pp. 1020 – 1033.

[14] Nedeljko Cvejic and Tapio Seppänen, Spread spectrum audio watermarking using frequency hopping and attack characterization, *Signal Processing*, Vol. 84, Issue 1, January 2004, pp. 207-213.

[15] Li, Lili; Hu, Jianling and Fang, Xiangzhong, Spread-Spectrum Audio Watermark Robust Against Pitch-Scale Modification, *Proc. of IEEE International Conference on Multimedia and Expo*, 2007, pp. 1770-1773.

[16] I. J. Cox, J. A. Bloom and M. L. Miller, *Digital Watermarking*, Morgan-Kaufman, 2002.

[17] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice-Hall, 2002.

[18] R. J. Anderson and F. A. P. Petitcolas, On the Limits of Steganography, *IEEE Journal of Selected Areas in Communication*, Vol. 16, No. 4, 1998, 474-481.

[19] A. Jazinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[20] A. Papoulis, *Signal Analysis*, McGraw-Hill, 1977.

[21] A. Bateman and W. Yates, *Digital Signal Processing Design*, Pitman, 1988.

[22] A. W Rihaczek, *Principles of High Resolution Radar*, McGraw-Hill, 1969.

[23] J. J. Kovaly, *Synthetic Aperture Radar*, Artech, 1976.

[24] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

[25] http://sound.media.mit.edu/mpeg4/audio/sqam/

[26] P. Kabal, An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality, Technical Report, McGill University, version 2, 2003.