High Resilience Diverse Domain Multilevel Audio Watermarking with Adaptive Threshold

Jerrin Thomas Panachakel and Anurenjan P.R.

Abstract A novel diverse domain (DCT-SVD & DWT-SVD) watermarking scheme is proposed in this paper. Here, the watermark is embedded simultaneously onto the two domains. It is shown that an audio signal watermarked using this scheme has better subjective and objective quality when compared with other watermarking schemes. Also proposed are two novel watermark detection algorithms viz., AOT (Adaptively Optimised Threshold) and AOTx (AOT eXtended). The fundamental idea behind both is finding an optimum threshold for detecting a known character embedded along with the actual watermarks in a known location, with the constraint that the Bit Error Rate (BER) is minimum. This optimum threshold is used for detecting the other characters in the watermarks. This approach is shown to make the watermarking scheme less susceptible to various signal processing attacks, thus making the watermarks more robust.

1 Introduction

The earliest reference in literature about audio watermarking is a patent titled, “Identification of sound and like signals,” filed by Emil Hembrooke of the Muzac Corporation in the year 1954 [1],[2]. Ever since then, many watermarking algorithms have been developed and deployed. Interestingly, the field of application of watermark-
Watermarking itself witnessed a gradual change. Today, watermarking finds wider applications such as transaction tracking, copyright protection, access control, broadcast monitoring etc. A good description of these applications can be found in [3]. Recently, watermarking has been used for patient identification, tampering detection etc. in wirelessly transmitted biomedical data such as ECG [4]. All these applications require the watermark to have one or more of the following properties,

1. **Robustness:** A watermark is said to be robust if it can survive common signal processing operations (commonly referred contextually as *signal processing attacks*) such as compression, re-sampling etc. [3]. A resourceful pirate can defeat a watermarking scheme by making it impossible to detect a watermark or making the scheme unreliable (i.e., a watermark is detected when no watermark is actually embedded) [5] both of which should be prevented. Some applications do not require the watermark to be robust and the watermarking schemes used in these applications are referred to as *fragile watermarking schemes* [6].

2. **Tamper Resistance:** Tamper resistance refers to the watermark’s ability to resist hostile attacks such as,
   - *Passive attacks* where the hacker tries to determine whether a watermark is present i.e. is trying to identify a covert communication.
   - *Active attacks* where the hacker tries to remove the watermark.
   - *Forgery attack* where the hacker tries to embed a new watermark [7].

Collusion attack is a special type of active attack where the hacker uses several copies of a work to produce a copy with no watermark [8].

3. **Fidelity:** A watermarking technique is said to have high fidelity if it causes only imperceptible degradation on the host signal. For instance, if the host signal is an audio data, it should be difficult to distinguish by hearing the original audio and the audio signal with the watermark embedded. This criterion is often referred to as the *perpetual transparency requirement*.

4. **Payload:** Payload refers to the amount of information that can be embedded into the given host signal. We require high payload while maintaining high perceptual transparency.

5. **Erasability:** In applications such as watermarking ECG signals in wireless transmission of biomedical signal, we require that the original host signal be recovered from the watermark signal.

Many watermarking schemes can be found in literature. These schemes can be broadly classified into two groups,

1. Transform domain watermarking where the watermark is embedded into the coefficients of the host in a transformed domain, like Discrete Wavelet Transform (DWT) [9], [10], [11], [12], Discrete Cosine Transform (DCT), [13], [14], Discrete Fourier Transform (DFT) [15], [16], [17], [18], DWT-SVD (Singular Value Decomposition) [19], etc., and

2. Spatial mode watermarking where embedding is done in the spatial domain itself, like the schemes given in [20] and [21].
Based on the host signal, the watermarking can be audio watermarking, image watermarking, biomedical signal watermarking etc.

Depending on the applications, the requirements on the watermarking scheme varies. This paper proposes a watermarking scheme in which one watermark is embedded in the DWT-SVD domain and the other watermark in the DCT-SVD domain. The watermarking in the DWT-SVD domain follows [19] but a novel watermarking scheme is followed in the DCT-SVD domain. Two different domains need to be used to avoid the interference between the two watermarks [22]. Also, two novel generic watermark extraction algorithms viz., AOT (Adaptively Optimised Threshold) and AOTx (AOT eXtended) are also proposed, which can be combined with any watermarking scheme that uses the ratio between a derived value (value obtained after applying transformation, decompositions, etc.) of the watermarked signal and the key (which is again a derived value but of the original host signal) as a threshold for extracting the watermark. It is shown that for the same payloads, the proposed diverse domain watermarking schemes outperforms conventional watermarking schemes in terms of imperceptibility. Also, it is shown that the use of AOT and AOTx at the detector side improves the robustness of the algorithm to various signal processing operations.

The rest of this paper is organised as follows, Section 2 gives an overview of the watermarking scheme; Subsections 2.2 and 2.3 discusses the embedding and extraction algorithms respectively. The performance metrics used for comparison are discussed briefly in Section 3. Finally, the results of the comparison of the proposed watermarking scheme with other watermarking schemes in terms of Bit Error Rate (BER) when the transmitted signal is corrupted by varying amounts of Additive White Gaussian Noise (AWGN), Mean Opinion Score (MOS) criterion for different genres and BER when subjected to various signal processing attacks are given in Section 4.

2 Proposed Watermarking Scheme

2.1 Overview

Most of the multilevel watermarking algorithms found in literature are for either image watermarking [23], [22] or for audio watermarking [24], [24], although a few works discuss about watermarking in other signals such as ECG signal [4]. M. Butman et al. propose a three level watermarking scheme in [22] in which spatial domain, DCT domain and wavelet domain are used for embedding watermarks in three levels. The primary purpose of this scheme is to detect tampering [22]. Contrary to this approach, S. Naveen et al. [24] use wavelet domain for multiple embedding with the primary purpose of increasing the payload. The primary disadvantage of the scheme by S. Naveen et al. is that the secondary and further watermarks can be recovered only under ideal conditions (noise free conditions) [24]. But in the
proposed watermarking scheme, the primary and secondary watermarks can be recovered even under non-ideal conditions. The use of Singular Value Decomposition (SVD) improves the imperceptibility (transparency) and robustness of the watermark [19], [24].

In almost all non-blind or informed watermarking schemes [3], the watermark extraction involves the comparison of the derived value of the embedded signal received at the receiver with a key generated at the embedding stage. But due to noise or some signal processing attacks, the derived value of the embedded signal obtained at the receiver may vary from its actual value i.e., the value just before the transmission stage (it is assumed that the corruption with noise, signal processing attacks, etc. occur in the transmission stage). This alters the ratio between the derived value and the key and this in turn will result in wrong decisions in the extraction stage. Adaptively Optimised Threshold (AOT) and AOTx (AOT eXtended) aim to minimise, if not eliminate completely, the error caused in this manner. The fundamental idea behind both AOT and AOTx is to embed a known character at a known location in the host signal, along with the original watermarks. At the receiver, the threshold of the detector is optimised so that BER of the detected watermark character at the location where the known character is embedded is minimum and this threshold is used for detecting the other characters in the watermarks. AOTx is superior to AOT in terms of the ability to determine the optimum threshold but has a cyclomatic complexity which is 6 times higher than that of the latter, thus making it computationally infeasible to apply AOTx at both levels.

2.2 Embedding Algorithm

![HLD of multilevel watermarking scheme.](image)

The high level diagram of the embedding process of the proposed watermarking scheme is given in Fig.1. Practically, there is only little significance in which watermarking is done first, whether it is the DCT-SVD or DWT-SVD, even though the watermarking is not strictly linear. In this work, DWT-SVD domain watermarking was done first followed by DCT-SVD domain watermarking. The two watermarking algorithms are discussed below.
2.2.1 DWT-SVD Domain

The watermarking scheme developed by Ali Al-Haj et al. described in [19] is used in this level, the only difference being that instead of binary image, the watermark is a text. Nevertheless, for the completeness of the discussion, the steps in the DWT-SVD watermarking scheme by Ali Al-Haj et al. are given below,

- **STEP 1:** The data to be watermarked, say $W$ is converted into its equivalent ASCII numbers and these ASCII numbers are converted to binary. Let the number of bits in the binary representation be $N$.
- **STEP 2:** The host audio signal is divided into $N$ frames. It is assumed that there are sufficient number of samples in the audio signal for this division. If the length of the audio signal, denoted as $L$ is not a multiple of $N$ sufficient zeroes are padded to the audio signal such that the new length $L'$ is a multiple of $N$.
- **STEP 3:** Four level DWT is performed on each frame to obtain the details sub-bands $D_1$, $D_2$, $D_3$ and $D_4$ and approximate sub-band $A_4$.
- **STEP 4:** Arrange the details sub-bands as shown in Fig. 2.

|  | $D_1$ | $D_2$ | $D_3$ | $D_4$ |
|---|---|---|---|---|
| $D_1$ | $D_2$ | $D_3$ | $D_4$ | $D_5$ |
| $D_6$ | $D_7$ | $D_8$ | $D_9$ | $D_{10}$ |
| $D_{11}$ | $D_{12}$ | $D_{13}$ | $D_{14}$ | $D_{15}$ |

Fig. 2 Matrix formulation of $X$.

- **STEP 5:** Perform SVD operation on $X$ to obtain the orthogonal matrix $U$, diagonal matrix $S$ and the unitary matrix $V$, all of which are real $4 \times 4$ matrices as given by the following equation,

$$SVD(X) = USV^T$$ (1)

- **STEP 6:** Let the non-zero elements in the $n^{th}$ diagonal matrix $S_n$, corresponding to the $n^{th}$ frame be denoted as $S_{11}$, $S_{22}$, $S_{33}$ and $S_{44}$. Modify the value of the diagonal element $S_{11}$ as follows,

$$S_{11} = S_{11} \times (1 + \alpha \times w(n))$$ (2)

where $\alpha$ is the watermark intensity, $w(n)$ is the $n^{th}$ watermark bit which is either 0 or 1. Higher the value of $\alpha$, more perceptible will the watermarking be. Lower values of $\alpha$ will make the detection more prone to errors. The original value of $S_{11}$, corresponding to each frame is stored as the “key”.
- **STEP 7:** Perform the inverse operations using the modified diagonal matrix to obtain the watermarked frame.
- **STEP 8:** Combine all frames to obtain the watermarked signal.

The high level diagram of DWT-SVD watermarking is given in Fig. 3.
2.2.2 DCT-SVD Domain

The embedding in the DCT-SVD domain follows a novel approach and is performed as follows:

- **STEP 1**: The data to be watermarked, say \( W \), is converted into its equivalent ASCII numbers and these ASCII numbers are converted to binary. Let the number of bits in the binary representation be \( N \).

- **STEP 2**: The host audio signal is divided into \( N \) frames. It is assumed that there are sufficient number of samples in the audio signal for this division. If the length of the audio signal, denoted as \( L \), is not a multiple of \( N \), sufficient zeroes are padded to the audio signal such that the new length \( L' \) is a multiple of \( N \).

- **STEP 3**: Apply DCT transformation on each frame.

- **STEP 4**: Form the matrix \( X \) for each frame by using the first three, middle three and last three elements of each transformed frame as shown in Fig.4.

\[
X = \begin{bmatrix}
1 & 4 & 7 \\
2 & 5 & 8 \\
3 & 6 & 9
\end{bmatrix}
\]

- **STEP 5**: Perform Singular Value Decomposition (SVD) operation on \( X \) to obtain the orthogonal matrix \( U \), diagonal matrix \( S \), and the unitary matrix \( V \), all of which are real \( 3 \times 3 \) matrices as given by the following equation,

\[
SVD(X) = USV^T
\]  (3)

- **STEP 6**: Let the non-zero elements in the \( n^{th} \) diagonal matrix \( S \) be denoted as \( S_{11}, S_{22} \) and \( S_{33} \). Modify the value of the diagonal element \( S_{11} \) as follows,
\[ S_{11} = S_{11} \times (1 + \alpha \times w(n)) \]  

(4)

where \( \alpha \) is the watermark intensity, \( w(n) \) is the \( n^{th} \) watermark bit which is either 0 or 1. The original value of \( S_{11} \) is stored as the “key”. Let the modified \( S \) matrix be denoted as \( S' \).

- **STEP 7:** Obtain matrix \( X' \) as,

\[ X' = US'V^T \]  

(5)

- **STEP 8:** \( X'' \) is obtained by applying inverse DCT on \( X' \). Apply the inverse of the operation performed in **STEP 4** to obtain the watermarked frame.

- **STEP 7:** Combine the \( N \) watermarked frames to obtain the watermarked audio signal.

The high level diagram of DCT-SVD watermarking is given in Fig. 5.

### 2.3 Extraction

For extraction, the inverse of the operations performed in the watermarking stage is performed. The watermark embedded last is extracted first and the one embedded first is extracted last. Though it may not be a necessary step in many watermarking schemes, here we require the audio signal onto which the embedding was done should be recovered at least after the first extraction since this proves critical in extracting the next watermark with least error. The algorithm for extracting the watermark if a static threshold (threshold determined during the embedding stage) is used is as follows;

Perform **STEPS 1-5** described in Subsection 2.2.1 and in Subsection 2.2.2 for extracting the watermark embedded in the DWT-SVD domain and DCT-SVD do-
main respectively. Let the first element in the diagonal matrix obtained after applying Singular Value Decomposition (SVD) be denoted as $S'_11$, and the same obtained in case of the original audio signal be denoted as $S_{11}$. It may be noted that it was these original values that were stored as the “key” in both domains. Theoretically, we detect the watermarked bit using the following rule,

$$\text{if } \left( \frac{S'_11}{S_{11}} \right) == \text{th}, \text{ then } w(n) = 1$$

$$\text{else } w(n) = 0$$

where $w(n)$ is the $n^{th}$ watermarked binary bit and $\text{th}$ is the threshold which is equal to $(1 + \alpha)$ where $\alpha$ is the watermark intensity. Practically, we do not go for a stringent condition as above, instead, we give an allowance to accommodate errors by replacing the equality condition with a comparative condition like,

$$\text{if } \left( \frac{S'_11}{S_{11}} \geq \text{th}' \right), \text{ then } w(n) = 1$$

$$\text{else } w(n) = 0$$

where $\text{th}'$ may be less than $\text{th}$ by a fraction of the value of $\text{th}$.

Now, when the transmitted signal gets modified due to some reasons such as addition of noise, signal processing operations, etc., the value of $S'11$ varies. Unless and otherwise a similar change in incorporated in either the “key” or the “threshold”, the detections in the receiver side can go wrong. $\text{AOT}$ and $\text{AOTx}$ modifies the “threshold” to what we refer to as the optimum threshold so that the detection errors are minimised.

$\text{AOT}$ is used for DCT-SVD watermark extraction and $\text{AOTx}$ is used for DWT-SVD watermark extraction. Once the optimum threshold is determined for each domain, these optimum thresholds are used for detecting the other characters embedded in the host signal in the respective domains. How $\text{AOT}$ and $\text{AOTx}$ are used for obtaining the optimum thresholds is as follows.

### 2.3.1 Adaptively Optimised Threshold (AOT)

Both $\text{AOT}$ and $\text{AOTx}$ are used for determining the optimum threshold that will minimise the BER at the detector. $\text{AOT}$ and $\text{AOTx}$ requires a known character to be embedded in a known location. The character “U” is chosen as the “known character” because the ASCII of “U” is 85, which in binary notation is 1010101, i.e., it has ‘1’s and ‘0’s in alternate locations, making the software programming easier. Also, it is the best go for a character having equal number of ‘1’s and ‘0’s in the binary of its ASCII equivalent.

Consider an audio signal in which a watermark in embedded along with the character “U” which is embedded as the first character in the watermark. Perform STEPS 1 to 5 given in Subsection 2.2.2. Let the non-zero elements in the $n^{th}$ diagonal ma-
trix $S_n$ be denoted as $S_{11}$, $S_{22}$ and $S_{33}$. Also, let the $n^{th}$ watermark bit be denoted as $w(n)$. Note that in this case, $w(0) = 1, w(1) = 0, w(2) = 1, w(3) = 0, w(4) = 1, w(5) = 0$ and $w(6) = 1$. Let $k(n)$ denote the original value of the top left element in the diagonal matrix obtained by performing SVD on the DCT applied frame, which can be obtained from the “key”. We define the following 6 thresholds,

$$thd_0 = \frac{w(0)}{k(0)}$$  \hspace{1cm} (6)  
$$thd_1 = \frac{w(1)}{k(1)}$$  \hspace{1cm} (7)  
$$thd_2 = \frac{w(2)}{k(2)}$$  \hspace{1cm} (8)  
$$thd_3 = \frac{w(3)}{k(3)}$$  \hspace{1cm} (9)  
$$thd_4 = \frac{w(4)}{k(4)}$$  \hspace{1cm} (10)  
$$thd_5 = \frac{w(5)}{k(5)}$$  \hspace{1cm} (11)  
$$thd_6 = \frac{w(6)}{k(6)}$$  \hspace{1cm} (12)  

Note that $thd_0, thd_2, thd_4$ and $thd_6$ correspond to bit 1 and $thd_1, thd_3$ and $thd_5$ corresponds to bit 0. The optimum threshold, $th_{opt}$ is given by,

$$th_{opt} = \max(th_0) + \frac{\min(th_1) - \max(th_0)}{2}$$  \hspace{1cm} (13)  

where $th_0$ is a vector formed using $thd_1, thd_3 & thd_5$ and $th_1$ is a vector formed using $thd_0, thd_2, thd_4$ and $thd_6$. Determination of optimum threshold $th_{opt}$ is shown graphically in Fig. 6.

![Fig. 6 Optimum threshold in AOT](image)

2.3.2 Adaptively Optimised Threshold eXtended (AOTx)

AOTx has 3 stages; the first stage is exactly similar to AOT. The threshold obtained from this stage is denoted as $th$. We define two variables $ze$ and $oe$. $ze$ is the ratio of number of ‘1’ s detected as ‘0’ s to the total number of ‘1’ s and $oe$ is the ratio of number of ‘0’ s detected as ‘1’ s to the total number of ‘0’ s. Clearly, the BER is a linear combination of $ze$ and $oe$. For the character “U”, $ze$, $oe$ and SNR are related as,
\[ SNR = \frac{oe \times 4 + ze \times 3}{7} \] (14)

What we are trying to achieve in AOTx is to minimise the sum of \( ze \) and \( oe \); i.e. we are actually minimising the BER.

If \( ze \) is greater than \( oe \), it means that the threshold is set at a lower value and therefore should be raised. Similarly, if \( oe \) is greater than \( ze \), it means that the threshold is set at a higher value and therefore should be lowered. If both \( ze \) and \( oe \) are 0, it means that \( th \) is the optimum threshold and there is no need for further optimisation.

In the actual implementation, the updating rule is as follows:

\[
\begin{align*}
&\text{if } (ze > oe), \text{ then } th = th + \frac{th}{2} \\
&\text{if } (ze == 0 \text{ and } oe == 0), \text{ then } th_{opt} = th; \text{ flag } = 1; \\
&\text{else } th = th - \frac{th}{2}
\end{align*}
\]

This is shown graphically in Fig. 7. flag is set when the optimum threshold is identified so that the final level of optimisation can be skipped to improve the computational time. In level 2, the change in threshold is rather coarse and during the implementation, we encountered situations where the thresholds determined in each subsequent iteration were oscillating between two values and the actual optimum threshold was between these two values. It is this observation that motivated us to design a third level in AOTx.

If the flag is set in level 2, level 3 is skipped. If it isn’t set, the updating rule is similar to that of level 2 except that instead of having a coarse variation, we now have a fine variation as shown below,
if (flag = 0), then thopt = th; break
if (ze > oe), then th = th + \frac{th}{10}
else th = th - \frac{th}{10}

3 Performance Metrics

The performance of the proposed watermarking scheme is compared with other watermarking techniques viz. DWT-SVD watermarking [19], simple multilevel watermarking with static threshold, multilevel watermarking with \textit{AOT} in DCT-SVD domain and \textit{AOTx} in DWT-SVD domain, double embedding (multilevel) watermarking proposed by S. Naveen \textit{et al.} in [24] (referred to hereafter as “Double Embedding”) and DCT-SVD domain watermarking with static threshold (referred to hereafter as “DCT-SVD based watermarking”). The static thresholds are decided at the watermark embedding stage itself. It may be noted that the DWT-SVD domain watermarking proposed by Ali Al-Haj \textit{et al.} has been shown in [19] to surpass the STFT-SVD based watermarking proposed by H. Ozer \textit{et al.} [25] and DCT based watermarking [26] proposed by I. Cox \textit{et al.}.

Three criteria are used for performance analysis,

| MOS | Description            |
|-----|------------------------|
| 5   | Imperceptible          |
| 4   | Perceptible but not annoying |
| 3   | Slightly annoying      |
| 2   | Annoying               |
| 1   | Very annoying          |

1. Mean Opinion Score (\textit{MOS}): \textit{MOS} gives a numerical value for the perceived quality of a signal. \textit{MOS} was used in this work to analyse the perceivable difference of the audio quality, i.e. it is a measure of the transparency of the watermarking algorithm. 5 scale \textit{MOS} is usually used with 5 being the best value and 1 being the worst value. The \textit{MOS} grading scale is given in Table I. Conventionally, \textit{MOS} is calculated by conducting listening test [19]. Contrary to this, we used “AQuA - Audio Quality Analyzer” developed by Sevana Oy (a Finnish limited company founded in 2003, providing software development and services including a voice quality assessment software) for performing audio quality tests.

2. BIT Error Rate (\textit{BER}): \textit{BER} was used to evaluate the watermark detection accuracy when the host signal is subjected to signal processing attacks, corrupted by noise etc. \textit{BER} is defined as [27], [19].
BER(W, \bar{W}) = \frac{\sum_{i=1}^{sz} W(i) \oplus \bar{W}(i)}{sz} \quad (15)

For the watermarking schemes that don’t support two simultaneous watermarks, the two watermarks are concatenated and \( W \) denotes the concatenated watermark and \( \bar{W} \) denotes the watermark recovered at the detector. \( sz \) corresponds to the size of the concatenated watermark, i.e. \( W \). For the watermarking schemes that support multiple watermarks, \( BER \) corresponding to both watermarks are calculated using the method described above and the average of the two is taken as the actual \( BER \).

3. **Signal to Noise Ratio (SNR):** \( SNR \) is used to measure the objective quality of the watermarked signal [28]. According to the recommendations of International Federation on Phonographic Industry (IFPI), the \( SNR \) of a watermarked audio signal should be at least 20\( dB \).

\( SNR \) is formulated in [27] as:

\[
SNR(X, \bar{X}) = 10\log_{10} \frac{\sum_{i=1}^{L} X(i)^2}{\sum_{i=1}^{L} (X(i) - \bar{X}(i))^2} dB \quad (16)
\]

where \( L \) is the number of samples in the original host signal \( X(i) \) and the watermarked signal \( \bar{X}(i) \).

4 Results

| Type of Attack | Watermarking Scheme | Double Embedding | Multilevel | Multilevel with AOT | DCT-SVD Domain | DWT-SVD Domain |
|---------------|---------------------|-----------------|------------|-------------------|----------------|----------------|
| Power Supply Hum | 0.00 | 0.00 | 0.00 | 0.36 | 0.36 |
| Amplification | 0.21 | 0.42 | 0.00 | 0.42 | 0.42 |
| Delay | 0.21 | 0.14 | 0.14 | 0.42 | 0.42 |
| Inversion | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Linear Transform and Sparsification | 0.35 | 0.14 | 0.00 | 0.35 | 0.50 |

The proposed watermarking scheme was simulated in MATLAB® 2012a (7.14). The waveform of the original host audio signal, which was of instrumental genre and the watermarked audio signal are given in Fig.4. The watermark text in the DCT-
SVD domain was “Jerrin Thomas Panachakel” and that in the DWT-SVD domain was “College of Engineering, Trivandrum, India”. There were 24 characters in the first watermark and the number of binary bits was 168. Similarly, the number of characters in the second watermark was 41 and the number of binary bits was 287. The character ‘U’ was embedded in the first frame of both domains for facilitating the use of $AOT$ and $AOT_x$. The watermark intensity used was $\alpha = 0.05$. 

Fig. 8 (a) Waveform of original host audio signal (b) Waveform of the watermarked audio signal
4.1 Bit Error Rate (BER)

4.1.1 AWGN Channel

Watermarking was performed on several genres on audio signal and this audio signal was corrupted with Additive White Gaussian Noise (AWGN) with varying Signal to Noise Ratio (SNR). The SNR was varied linearly from 0 dB to 100 dB in steps of 10 dB. This is assumed to model an AWGN channel through which the signal may pass during transmission. This assumption is supported by the central limit theorem and hence can be considered fair. Plots between BER and SNR for various genres are given in Fig. 9 and Fig. 10. It is evident from the figure that Multilevel Watermarking with AOT has the lowest BER when the watermark signal is corrupted by AWGN noise. The plots are not monotonic in nature due to the randomness associated with the addition of noise.
Fig. 10 BER v/s SNR for different genres. (a) Ballad (b) Pop Rock (c) Speech
4.2 Signal processing attacks

The host signal was watermarked using various watermarking schemes with two random watermarks if the scheme supports multiple watermarks or with the watermark obtained by concatenating the two random watermarks if the scheme supports only a single watermark. The watermarked signals were subjected to various signal processing attacks such as:

- **Power supply hum**: A 0.25V_{pp} sine wave of 50Hz is added to the watermarked audio signals to simulate the effect of addition of power supply hum. The corrupted watermarked signal was then processed for recovering the watermarks.
- **Amplification**: Each of the watermarked signals is amplified by 14dB.
- **Delay**: The watermarked signals are delayed by 100ms.
- **Inversion**: All the samples in the watermarked signals are inverted.
- **Linear Transformation and Sparsification**: Discrete Cosine Transform (DCT) was applied on the watermarked signals and the DCT coefficients whose absolute magnitude was less than 0.05 was made zero. In our work, a speech signal was used as the host signal and 68% of the DCT coefficients were less than 0.05 and was hence made zero. These operations were done to model the effects of a lossy compression.

After applying the above signal processing operations, the resultant watermarked signal was processed for recovering the watermark signal. Contrary to the approach by Naveen *et al.* in [24] where filtering was done prior to extraction, we attempted the recovery of the watermark without filtering the corrupted watermarked signal. The BER between the original signal and the watermarked signal of various watermarking schemes when subjected to the above signal processing attacks is given in Table II. In all the cases, the proposed watermarking scheme has comparative, if not better performance, i.e. lowest BER.

4.3 Mean Opinion Score (MOS) and Signal to Noise Ratio (SNR)

The MOS and SNR values were calculated for different watermark intensities \( \alpha = 0.05 \) & \( \alpha = 0.1 \) for a speech signal sampled at 8KHz and is shown in Table III & IV. Clearly, the proposed technique has better MOS values, proving our argument that watermarking on two diverse domains has better subjective quality than watermarking on a single domain. The values obtained for SNR satisfies the International Federation on Phonographic Industry (IFPI) recommendations. It is interesting to note that although the DCT-SVD domain based watermarking is vulnerable to signal processing attacks, it has an SNR greater than the value specified by IFPI for a watermarked audio signal, 60.24dB and 54.11dB for \( \alpha \) values 0.05 and 0.1 respectively. These values are highest when compared to the values of other watermarking schemes discussed. Also, the MOS values for DCT-SVD domain watermarking is comparable to multilevel watermarking.
Table 3  MOS and SNR values for $\alpha = 0.05$

| Watermarking Technique | MOS  | SNR (dB) |
|------------------------|------|----------|
| Double Embedding       | 4.55 | 27.00    |
| Multilevel Watermarking| 4.58 | 33.20    |
| DCT-SVD                | 4.33 | 60.24    |
| DWT-SVD                | 4.33 | 30.45    |

Table 4  MOS and SNR values for $\alpha = 0.1$

| Watermarking Technique | MOS  | SNR (dB) |
|------------------------|------|----------|
| Double Embedding       | 4.37 | 20.76    |
| Multilevel Watermarking| 4.55 | 27.18    |
| DCT-SVD                | 4.26 | 54.11    |
| DWT-SVD                | 4.24 | 24.42    |

5 Conclusion

Novel multilevel watermarking in diverse domains, DCT-SVD domain based watermarking and two generic watermark extraction algorithms viz., Adaptively Optimised Threshold (AOT) and AOT eXtended (AOTx) are discussed. The proposed watermarking algorithm has better subjective and objective quality, which is evident from the higher SNR and MOS values. The use of AOT and AOTx at the detection stage makes the watermarking schemes less susceptible to various signal processing attacks although the proposed technique fails for geometric signal processing attacks such as cropping, addition of delay etc. This demerit can be addressed and resolved by using synchronisation codes. The fact that AOT and AOTx are generic algorithms which can be incorporated in several existing watermarking schemes to improve their susceptibility towards signal processing attacks opens up a wide range of applications for both. Also, the developed DCT-SVD domain based watermarking has the highest SNR eventhough it is more susceptible to signal processing attacks. Hence, this can be used as a “fragile watermarking scheme”.
Acknowledgement

We would like to place on record our deepest gratitude to

- Mr. Valeri Sitnikov of “Sevana Oy” for providing us a free evaluation version of their product, “AQuA - Audio Quality Analyzer” for this work.
References

[1] Cox, I.J., Miller, M.L.: The first 50 years of electronic watermarking. EURASIP J. Appl. Signal Process. 2002 (2002) 126–132
[2] Frank, H.E.: Identification of sound and like signals (1961)
[3] Cox, I.J., Miller, M.L., Bloom, J.A.: Watermarking applications and their properties. In: International Conference on Information Technology: Coding and Computing. (2000) 6–10
[4] Kaur, S., Farooq, O., Singhal, R., Ahuja, B.: Digital watermarking of ecg data for secure wireless communication. In: Recent Trends in Information, Telecommunication and Computing (ITC), 2010 International Conference on. (2010) 140–144
[5] Swanson, M.D., Zhu, B., Tewfik, A.H., Boney, L.: Robust audio watermarking using perceptual masking. Signal Process. 66 (1998) 337–355
[6] Cvejic, N.: Algorithms for Audio Watermarking and Steganography. Acta Universitatis Ouluensis: Technica. Oulun yliopisto (2004)
[7] Craver, S., Memon, N., Yeo, B.L., Yeung, M.: Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications. Selected Areas in Communications, IEEE Journal on 16 (1998) 573–586
[8] Cox, I.J., Kilian, J., Leighton, F., Shamoon, T.: Secure spread spectrum watermarking for multimedia. Image Processing, IEEE Transactions on 6 (1997) 1673–1687
[9] Podilchuk, C., Zeng, W.: Image-adaptive watermarking using visual models. IEEE J. Select. Areas Commun. 16 (1998) 525–539
[10] Kaeunmerd, N., Rao, K.: Wavelet based image adaptive watermarking scheme. Electronics Letters 36 (2000) 312–313
[11] Piva, A., Bartolini, F., Boccardi, L., Cappellini, V., De Rosa, A., Barni, M.: Watermarking through color image bands decorrelation. In: Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on. Volume 3. (2000) 1283–1286 vol.3
[12] Luo, L., Chen, Z., Chen, M., Zeng, X., Xiong, Z.: Reversible image watermarking using interpolation technique. Information Forensics and Security, IEEE Transactions on 5 (2010) 187–193
[13] Shih, F.Y., Wu, Y.T.: Decomposition of binary morphological structuring elements based on genetic algorithms. Computer Vision and Image Understanding 99 (2005) 291 – 302
[14] Wei, Z., Ngan, K.: Spatio-temporal just noticeable distortion profile for grey scale image/video in dct domain. Circuits and Systems for Video Technology, IEEE Transactions on 19 (2009) 337–346
[15] Tsui, T.K., Zhang, X.P., Androutsos, D.: Color image watermarking using multidimensional fourier transforms. Information Forensics and Security, IEEE Transactions on 3 (2008) 16–28
[16] Shih, F.Y.: Digital Watermarking and Steganography: Fundamentals and Techniques. 1st edn. CRC Press, Inc., Boca Raton, FL, USA (2007)
[17] Tsui, T.K., Zhang, X.P., Androutsos, D.: Quaternion image watermarking using the spatio-chromatic fourier coefficients analysis. In: Proceedings of the 14th annual ACM international conference on Multimedia. MULTIMEDIA ’06, New York, NY, USA, ACM (2006) 149–152
[18] Bas, P., le Bihan, N., Chassery, J.M.: Color image watermarking using quaternion fourier transform. In: Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP ’03). 2003 IEEE International Conference on. Volume 3. (2003) III–521–4 vol.3
[19] Al-Haj, A., Mohammad, A.: Digital audio watermarking based on the discrete wavelets transform and singular value decomposition. European Journal of Scientific Research 39 (2010) 6–21
[20] Kalker, T., Haitisma, J.: Efficient detection of a spatial spread-spectrum watermark in mpeg video streams. In: Image Processing. 2000. Proceedings. 2000 International Conference on. Volume 1. (2000) 434–437 vol.1
[21] Hernández, J., Rodríguez, J., Prez-González, F.: Improving the performance of spatial watermarking of images using channel coding. Signal Processing 80 (2000) 1261 – 1279
[22] Butman, M., Hel-Or, H.: Multi-level watermarking with independent decoding. In: Image Processing, 2001. Proceedings. 2001 International Conference on. Volume 2. (2001) 514–517 vol.2
[23] Shaikh, M., Dote, Y.: A robust multilevel watermarking method for digital images using multiresolution wavelet decomposition. In Roy, R., Kppen, M., Ovaska, S., Furuhashi, T., Hoffmann, F., eds.: Soft Computing and Industry. Springer London (2002) 363–370
[24] Naveen, S., Shiras, S., Nithin, N., Tony, J., Gayathri, K.: Multiple audio marking using discrete wavelets transform and singular value decomposition. In: Emerging Technologies, 2012. Proceedings. First National Conference on. Volume 1. (2012) 6–12
[25] Özer, H., Sankur, B., Memon, N.: An svd-based audio watermarking technique. In: Proceedings of the 7th workshop on Multimedia and security. MM&#38;Sec ’05, New York, NY, USA, ACM (2005) 51–56
[26] Cox, I.J., Kilian, J., Leighton, F., Shamoon, T.: Secure spread spectrum watermarking for multimedia. Image Processing, IEEE Transactions on 6 (1997) 1673–1687
[27] K, V.B., Sengupta, I., Das, A.: An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. Digital Signal Processing 20 (2010) 1547 – 1558

[28] Erelebi, E., Batak, L.: Audio watermarking scheme based on embedding strategy in low frequency components with a binary image. Digital Signal Processing 19 (2009) 265 – 277