Statistical Audio Watermarking Algorithm Based on Perceptual Analysis

Xiaomei Quan
Computer Institute, Beijing University of Techology
Pingleyuan 100#, Chaoyang District,
Beijing, China
8610-67392717
xmquan@emails.bjut.edu.cn

Hongbin Zhang
Computer Institute, Beijing University of Technology
Pingleyuan 100#, Chaoyang District,
Beijing, China
8610-67392878
zhb@public.bta.net.cn

ABSTRACT
In this paper, we describe a novel statistical audio watermarking scheme. Under the control of the masking thresholds, watermark is embedded adaptively and transparently in the perceptual significant portions in wavelet packet domain by a statistical method. Watermark detection can be done without access to the original signal. Experimental results show the proposed scheme can survive common signal manipulations and malicious attacks.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information System**]: Audio input/output, Evaluation/methodology; K.5.1 [**Legal Aspects of Computing**]: Hardware/Software Protection-Copyrights.

General Terms
Security

Keywords
Audio watermarking, psychoacoustic model, wavelet packet decomposition

1. INTRODUCTION
The proliferation of digital multimedia content and powerful PCs as well as broadband internet connections provide for public the conveniences to modify and redistribute digital content at will. However along with such conveniences come new risks for digital multimedia content security, including copyright violation, illegal content modification, and content tampering, to name a few. Digital watermarking as an effective deterrence to above illegal behaviors has developed quickly over the past decade. Digital watermarking is directly embedding some additional information, e.g. information of content owners or authors, into the original content or host signal imperceptibly. Apart from copyright protection, digital watermark can also serve as an additional tool for content authentication, copy control, broadcast monitoring and secret communication, etc. Watermarking for different application should meet different requirements [4, 12]. In this paper, we only focus on the robust watermarking intended for copyright protection. The essential part of this kind of watermarking is to find the balance point between robustness and imperceptibility. To improve robustness while maintain imperceptibility has been an active research topic all the while [9, 10, 19, 20].

One effective way to achieve a trade-off between robustness and imperceptibility is to take advantage of the perceptual or statistical properties of host signals. But unfortunately, among currently published watermarking schemes, the statistical ones are relatively few, and few of them employ perceptual criteria to ensure transparency. The typical statistical watermark is patchwork scheme and its derivatives [4], [2, 22].

In this paper a novel statistical audio watermarking method in wavelet packet domain is proposed. The scheme improves watermark robustness while maintaining transparency by increasing the watermark strength within the limits of masking thresholds. The main contribution of this paper is threefold. Firstly, it incorporates perceptual analysis into statistical method and estimates masking thresholds in both time and wavelet packet domain, therefore the transparency of the proposed scheme is guaranteed objectively. It adapts wavelet packet decomposition to HAS, and calculates the masking thresholds of the host audio using MPEG psychoacoustic model I for layer I [8] in wavelet packet domain. Since wavelet packet decomposition has more flexible time/frequency resolution than FFT does, the produced subband structure is closer to that of Critical Band, which is decomposed by human ear. Thus psychoacoustic model in wavelet packet domain more precisely reflects how human ear actually works than the one in FFT domain [1, 13, 17, 18]. The perceptual analysis helps to determine the maximum perceptual redundancy in host audio. Secondly, this paper embeds watermark in tonal components, which makes the proposed method more robust, especially against low bit-rate encoding and time scaling modification attacks. Thirdly, comparing with the previous patchwork methods, this paper doubles the embedded data rate by mapping watermark to different embedding subsets makes it more suitable for practical applications.

In next section, we describe how to estimate masking thresholds. Watermarking embedding and detection algorithms are presented in Section 3. Performance analysis and experimental results are given in Section 4. Finally, we conclude this paper in Section 5.

DRM'05, November 7, 2005, Alexandria, Virginia, USA.

2. MASKING THRESHOLD ESTIMATION

The existence of auditory masking makes it possible to embed watermark in audio signal imperceptible. So it is necessary to integrate masking threshold into watermarking scheme to ensure its transparency objectively. Masking thresholds are estimated by psychoacoustic models. MPEG Psychoacoustic models [8] are most commonly used ones at present. But as noted in [13], they have inherent deficiencies, i.e., the subband division in MPEG cannot precisely reflects Critical Bands, therefore the masking thresholds derived from them are more conservative. Due to this reason, wavelet packet decomposition is introduced into psycho-acoustic model in some audio compression and watermarking algorithms [14, 18]. By virtue of wavelet packet's flexible decomposed subband structure and good temporal/frequency resolution, these algorithms can calculate masking threshold better. However, these algorithms still use FFT to calculate masking threshold, which adds extra computation burden on the model. In fact wavelet packet decomposition provide enough time and frequency resolution to estimate the thresholds without the help of FFT. So in this paper we only use wavelet packet decomposition to accomplish signal mapping from time to frequency domain and calculate masking threshold directly in wavelet packet domain [5]. In addition, we also incorporate temporal masking model to determine post-masking threshold, which is the most important masking threshold in time domain.

2.1 Frequency Masking Threshold

Masking thresholds calculation in wavelet packet domain starts with wavelet packet decomposition. Developed from wavelet transformation, which only decomposes approximation spaces, wavelet packet splits both approximation and detail spaces. Each space splitting corresponds to a kind of decomposition, identified with a dyadic tree called admissible tree, from whose leaf node, we get the wavelet packet basis. All wavelet packet bases constitute the wavelet packet library. The best basis selection is application-dependent. In this paper, the goal is to get a closer approximation to Critical Bands, so we use 8-stage, 29-band wavelet packet tree in [17], which is shown in Figure 1. Each node on an admissible tree is labeled by (k, p), where k denotes scale and p is the number of nodes that are on its left at the same depth. The leave nodes correspond to the decomposed subbands.

An example frame (512 samples, blues sampled at 44.1 kHz/16-bit resolution) and its wavelet packet decomposition result are shown in Figure 2 and Figure 3, respectively.

We use subband signal $S(z_i)$ to calculate the power density spectrum $X(z_i)$, where $S(z_i)$ is in subband (k, p) and z_i is the frequency on the bard scale. $X(z_i)$ is determined by:

$$X(z_i) = 10 * \log_{10}|S(z_i)|^2 \quad dB \tag{1}$$

Within each subban (k, p), the maximum power density spectrum is the sound pressure level in (k, p).

Since tonal and non-tonal components have different masking behaviors, they should be discriminated. Find the local maximum of the power density spectrum; if the difference between two adjacent maximum components is greater than 7 dB, label the local maximum as tonal component. Power density spectrum and tonal components on the example frame are illustrated in Figure 4.

The non-tonal components can be calculated from rest of the spectrum. Tonal and non-tonal components have different masking index av_{tm} and av_{nm}, but same masking function v_f.

$$\begin{cases} av_{tm} = -1.525 - 0.275 z_i - 4.5 dB \\ av_{nm} = -1.525 - 0.175 z_i - 0.5 dB \end{cases} \tag{2}$$

$$v_f(z_i, z_j) = \begin{cases} 17\Delta z - 0.4 X(z_i) + 11, & -3 \le \Delta z < -1 \\ (0.4 X(z_i) + 6)\Delta z, & -1 \le \Delta z < 0 \\ -17\Delta z, & 0 \le \Delta z < 1 \\ -17\Delta z + 0.15(\Delta z - 1) X(z_i), & 1 \le \Delta z < 8 \end{cases} \tag{3}$$

where $\Delta z = z_i - z_j$ is frequency separation.

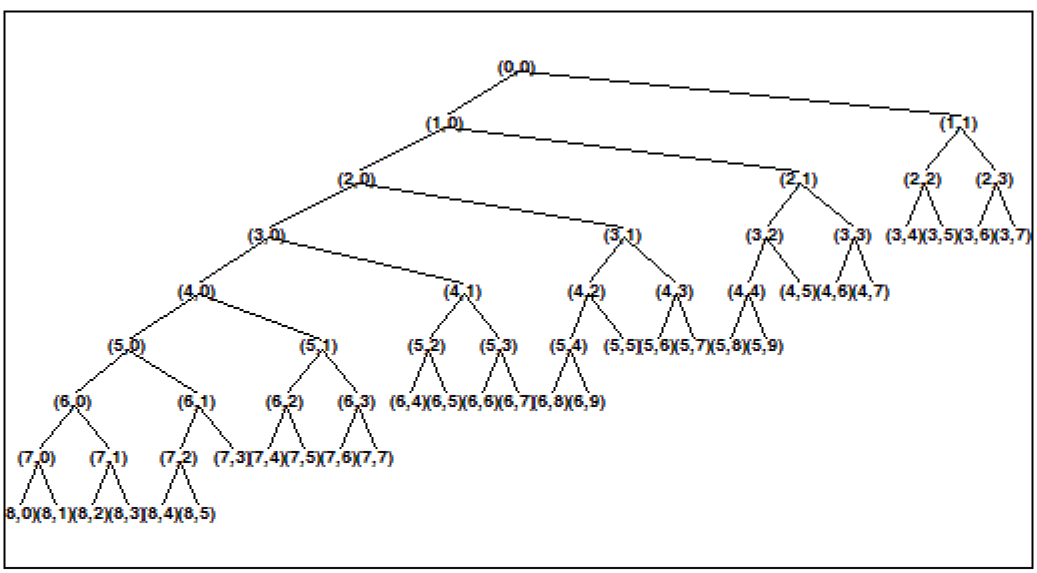

Figure 1. Wavelet packet decomposition tree.

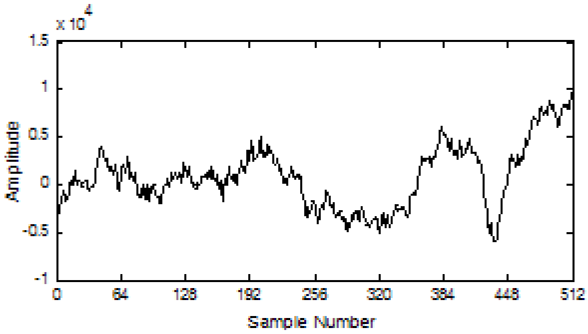

Figure 2. Example frame with 512 samples.

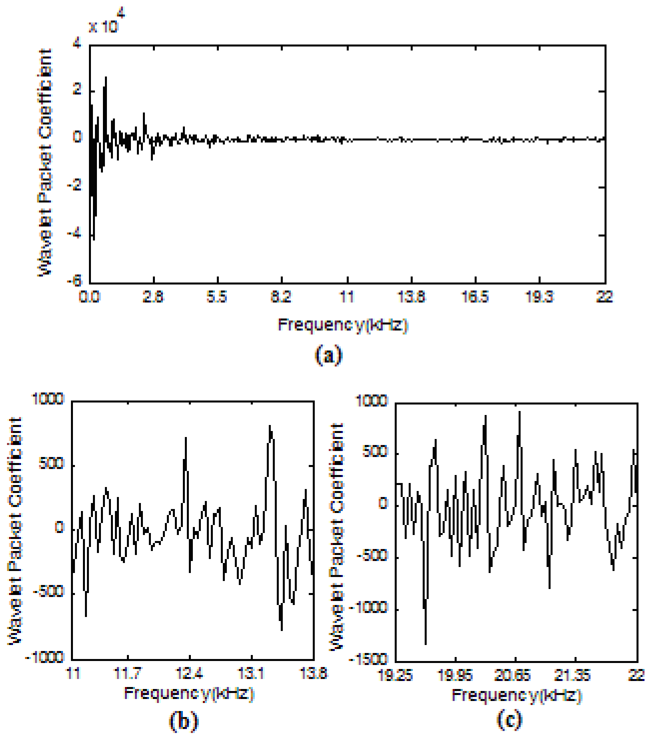

Figure 3. Wavelet packet decomposition result
(a) whole frame (b) subband (3,4) (c) subband(3,7).

With av_{tm} , av_{nm} as well as v_f , the individual masking thresholds of tonal and non-tonal components can be calculated by,

$$\begin{cases} LT_{tm}(z_i,z_j) = X_{tm}(z_j) + av_{tm} + v_f(z_i,z_j) \\ LT_{nm}(z_i,z_j) = X_{nm}(z_j) + av_{nm} + v_f(z_i,z_j) \end{cases} \quad (4)$$

Then the global threshold LT_g and the final masking threshold $LT_{\min}$ in subband (k, p) are:

$$LT_g(z_i) = 10\log_{10}(10^{LT_q(z_i)} + \sum_{j=1}^{N_t} 10^{LT_{tm}(z_i,z_j)/10} + \sum_{j=1}^{N_n} 10^{LT_{nm}(z_i,z_j)}) \quad (5)$$

$$LT_{\min}(k, p) = \min_{LT_g(z_i)\in(k,p)}[LT_g(z_i)] \quad (6)$$

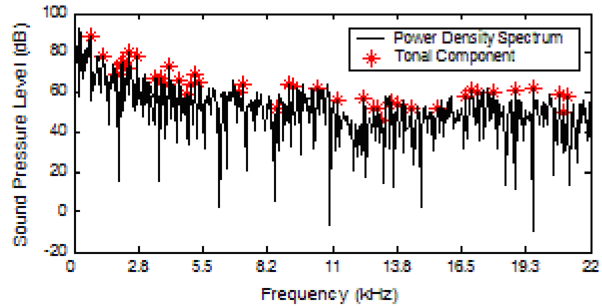

Figure 4. Power density spectrum and tonal components.

Figure 5 shows the final masking thresholds comparison between wavelet packet subbands and MPEG 32-equal-sized subbands.

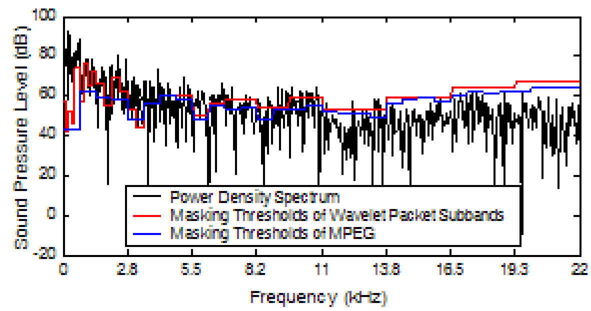

Figure 5. Masking thresholds comparison

2.2 Temporal Masking Threshold

We use the temporal masking threshold in [16] to calculate the post-masking threshold $TF(k, p)$ produced by subband (k, p).

$$TF(k, p) = 0.5(\log_{10} 200 - \log_{10} t)(10\log_{10} \frac{1}{2^k} \sum_{i=0}^{2^k-1} S^2(z_i) - \alpha) \quad (7)$$

where t is the time distance between masker and maskee, and parameter $\alpha = (-0.1123f^2 + 2.979f + 6.6727)$, f is center frequency of subband (k, p).

Suppose the duration of the post-masking is 200ms, the sampling frequency is 44.1kHz, then in each subband the post-masking threshold $TM(k, p)$ is the maximum of the previously 18 successive frames. Based on power law, overall masking threshold in subband is $MT(k, p)$:

$$MT(k, p) = (LT^{\lambda}(k, p) + TM^{\lambda}(k, p))^{\frac{1}{\lambda}} \quad (8)$$

where $\lambda = 1/3$.

3. WATERMARKING EMBEDDING AND DETECTION ALGORITHMS

3.1 Embedding Algorithm

We use a statistical method to embed two-bit watermark each time, which is motivated by patchwork algorithm. We select perceptual significant portions of host signal, namely tonal component in wavelet packet domain, to embed watermark.

Suppose one watermark dataset contains N tonal components, the embedding procedure is as follows:

1) From N tonal components, use the secret key and two-bit watermark, 00, 01, 10, 11, to pseudo-randomly select two subsets: $A_s|_{s=0,1} = \{a_{st}\}_{t=1,\dots,n}$ and $B_s|_{s=0,1} = \{b_{st}\}_{t=1,\dots,n}$, where $2n \le N$. The subscripts of A and B correspond to the embedded two-bit watermark, which is illustrated in Figure 6. For instance, we select subsets A_0 and B_1 if we intend to embed watermark 01.

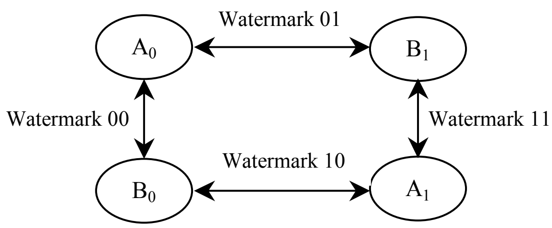

Figure 6. Subsets selection rule

2) Assume MT_a and MT_b are the overall subband masking thresholds corresponding to selected subset A and B. the embedding function is,

$$\begin{cases} a' = a + MT_a * \eta \\ b' = b - MT_b * \eta \end{cases} \tag{9}$$

where $\eta \in (0,1)$ and adapts to the statistical and auditory properties of the host signal. If the masking threshold is less than the sample mean, that means redundancy in this component is relatively few, the watermark strength is proportional to the ratio between masking threshold and the sample mean. Otherwise if there is much redundancy in the component, we adjust η close to its maximum.

$$\eta = \begin{cases} \dfrac{MT_a + MT_b}{\bar{a} + \bar{b}} & MT_a + MT_b < \bar{a} + \bar{b} \\ 0.95 & MT_a + MT_b \ge \bar{a} + \bar{b} \end{cases} \tag{10}$$

3) Replace the original coefficient with watermarked ones, and apply inverse wavelet packet transformation. Since η is less than 1, artifacts introduced by watermark embedding is bounded by masking threshold MT, which renders the whole embedding process inaudible.

3.2 Detection Algorithm

Watermark detection is based on the comparison of the test statistics on the candidate subsets A and B. The detection process is as follows:

1) Carry out the same wavelet packet decomposition on the audio frame, and find the watermark dataset.

2) Map the secret key and watermark to the seed of the random number generator to generate the subsets A_0, A_1, B_0 and B_1. Then calculate the sample mean, $\bar{a}_i|_{i=0,1}$ $\bar{b}_j|_{j=0,1}$ and the sample variance $S^2_{A_i}|_{i=0,1}$ $S^2_{B_j}|_{j=0,1}$ on the above candidate subsets, respectively:

$$\bar{a}_i = \frac{1}{n}\sum_{t=1}^{n} a_{it} \quad \bar{b}_j = \frac{1}{n}\sum_{t=1}^{n} b_{jt}$$

$$S^2_{A_i} = \frac{1}{n-1}\sum_{t=1}^{n}(a_{it} - \bar{a}_i)^2 \quad S^2_{B_j} = \frac{1}{n-1}\sum_{t=1}^{n}(b_{jt} - \bar{b}_j)^2$$

3) Calculate the test statistic:

$$T_{ij} = \frac{\sqrt{\frac{n}{2}}(\bar{a}_i - \bar{b}_j)}{\sqrt{\frac{(n-1)S^2_{A_i} + (n-1)S^2_{B_j}}{2(n-1)}}}$$

Let $T_{pq} = \max[T_{ij}]_{i,j=0,1}$, compare T_{pq} with the pre-set threshold T_m, if $T_{pq} > T_m$, two watermark bits are supposed to be embedded in the subset A_p and B_q. The extracted watermark bits will be pq. Otherwise, we assume there isn't any embedded watermark in subsets A_0, A_1, B_0 and B_1.

In order to decide the probability of various watermark detection error and quantitatively analyze the robustness of the proposed watermark method, we set up the statistical hypothesis test. Two competing hypotheses are the null hypothesis, denoted as H_0, against the alternative hypothesis, denoted as H_1.

H_0: Two-bit watermark is not embedded;

H_1: Two-bit watermark is embedded.

When watermark is embedded, the probability density function (pdf) of the test statistic T is $\varphi_1(t)$, otherwise, the pdf is $\varphi_0(t)$.

Assume the component of the watermark dataset is normally distributed with mean μ and variance σ^2. Let $\delta = MT * \eta$, then,

1) Under H_0:

$$a_i|_{i=0,1} \sim N(\mu, \sigma^2), \; b_i|_{i=0,1} \sim N(\mu, \sigma^2) \text{ and}$$

$$\bar{a}_i|_{i=0,1} \sim N(\mu, \frac{\sigma^2}{n}), \; \bar{b}_j|_{j=0,1} \sim N(\mu, \frac{\sigma^2}{n}),$$

$$\bar{a}_i - \bar{b}_j \sim N(0, \frac{2\sigma^2}{n})$$

Therefore $T \sim \varphi_0(t) = t(2n-2)$, where $t(2n-2)$ denotes Student's t-distribution with *2n-2* degrees of freedom.

2) Under H_1:

$$a_i|_{i=0,1} \sim N(\mu + \delta, \sigma^2), \; b_i|_{i=0,1} \sim N(\mu - \delta, \sigma^2) \text{ and}$$

$$\bar{a}_i|_{i=0,1} \sim N(\mu + \delta, \frac{\sigma^2}{n}), \; \bar{b}_j|_{j=0,1} \sim N(\mu - \delta, \frac{\sigma^2}{n})$$

$$\overline{a_i} - \overline{b_j} \sim N(2\delta, \frac{2\sigma^2}{n}),$$

Therefore $T \sim \varphi_1(t) = t(2n-2, 2\delta)$, and $t(2n-2, 2\delta)$ denotes the noncentral Student's t-distribution with $2n-2$ degrees of freedom and noncentrality parameter 2δ.

With $\varphi_0(t)$ and $\varphi_1(t)$, the type I error (H_0 is rejected when in fact the watermark is not embedded) and type II error (H_0 is not rejected when in fact the watermark is embedded) as well as the bit error rate (BER) can be calculated.

1) Type I error P_I:

Since $P(T_{pq} = \max[T_{ij}]_{i,j=0,1}) = \frac{1}{4}$, each time two-bit watermark pq is embedded, the type I error α is given by,

$$\begin{aligned}\alpha &= P(T_{pq} > T_m \mid H_0) \\ &= P(T_{pq} = \max[T_{ij}]_{i,j=0,1}) \int_{T_m}^{+\infty} \varphi_0(t, 2n-2) dt \\ &= \frac{1}{4} \int_{T_m}^{+\infty} \varphi_0(t, 2n-2) dt\end{aligned}$$

where

$$\varphi_0(t, 2n-2) = \left[\Gamma(\frac{2n-1}{2}) \Big/ \sqrt{(2n-2)\pi}\, \Gamma(n-1)\right] \left(1 + \frac{t^2}{2n-2}\right)^{-\frac{2n-1}{2}}$$

and $\Gamma(z)$ is the gamma function.

Every two-bit watermark pq is repeatedly embedded R times using the above proposed scheme. Obviously, they are R Bernoulli trials. Define X to be the number of events that $T_{pq} > T_m$ obtained in R trials. Only if X is not less than half times of R, the two-bit watermark pq is believed to be extracted successfully. Therefore, the overall type I error can be decided by:

$$P_I = P(X \geq \frac{R}{2}) = \sum_{x=\frac{R}{2}}^{R} \binom{R}{x} \alpha^x (1-\alpha)^{R-x}$$

2) Type II error P_{II}:

The type II error β on each trial is calculated by,

$$\begin{aligned}\beta &= P(T_{pq} \leq T_m \mid H_1) \\ &= P(T_{pq} = \max[T_{ij}]_{i,j=0,1}) \int_{-\infty}^{T_m} \varphi_1(t, 2n-2, 2\delta) dt \\ &= \frac{1}{4} \int_{-\infty}^{T_m} \varphi_1(t, 2n-2, 2\delta) dt\end{aligned}$$

where

$$\varphi_1(t, n, \lambda) = \frac{n^{n/2} n!}{2^n e^{\lambda^2/2} (n+t^2)^{n/2} \Gamma(\frac{n}{2})} \cdot \left\{ \frac{\sqrt{2}\lambda t_1 F_1(\frac{n}{2}+1; \frac{3}{2}; \frac{\lambda^2 t^2}{2(n+t^2)})}{(n+t^2)\Gamma\left[\frac{1}{2}(n+1)\right]} + \frac{{}_1F_1\left(\frac{1}{2}(n+1); \frac{1}{2}; \frac{\lambda^2 t^2}{2(n+t^2)}\right)}{\sqrt{n+t^2}\,\Gamma(\frac{1}{2}n+1)} \right\}$$

where $\Gamma(z)$ is the gamma function and ${}_1F_1(a;b;z)$ is a confluent hypergeometric function,

$${}_1F_1(a;b;z) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{zt} t^{a-1} (1-t)^{b-a-1} dt$$

Then the overall type II error is decided by,

$$P_{II} = P(X < \frac{R}{2}) = \sum_{x=0}^{\frac{R}{2}-1} \binom{R}{x} \beta^x (1-\beta)^{R-x}$$

3) Bit error rate (BER) P_{BER}:

Bit error rate refers to the probability that the extracted two-bit watermark isn't the same as the embedded watermark. For instance, when the extracted watermark is 01, 10, or 11 while the embedded watermark is 00, bit error occurs. In this case, we denote the BER as γ_{01}. Similar BERs are denoted as γ_{01}, γ_{10} and γ_{11}. According to the above proposed embedding/detection scheme and the definition of BER, it can be easily understood that all the BERs, γ_{00}, γ_{01}, γ_{10} and γ_{11}, are equal, i.e. $\gamma_{00} = \gamma_{01} = \gamma_{10} = \gamma_{11} = \gamma$.

Let γ_{00-01} be the BER that "00" is marked but "01" is detected. Suppose T_{00-01} is the dividing point of the two pdf φ_{00} and φ_{01}, then γ_{00-01} can be obtained by

$$\gamma_{00-01} = P(T_{01} > T_{00}, T_{01} > T_m \mid embedded 00) = \frac{1}{3} \int_{-\infty}^{T_{00-01}} \varphi_{00}(t) dt$$

Similarly $\gamma_{00-10} = \frac{1}{3} \int_{-\infty}^{T_{00-10}} \varphi_{00}(t) dt$ and $\gamma_{00-11} = \frac{1}{3} \int_{-\infty}^{T_{00-11}} \varphi_{00}(t) dt$.

Hence $\gamma_{00} = \gamma_{00-01} + \gamma_{00-10} + \gamma_{00-11}$. Therefore on each trial, the detection BER can be decided by $\gamma = \gamma_{00}$. So the total BER on R trials can be calculated by

$$P_{BER} = P(X \geq \frac{R}{2}) = \sum_{x=\frac{R}{2}}^{R} \binom{R}{x} \gamma^x (1-\gamma)^{R=x}$$

4. PERFORMANCE ANALYSIS AND EXPERIMENTAL RESULTS

By adopting the following measures, the performance of the proposed watermarking algorithm is improved.

1) To map two-bit watermark to watermark subsets, watermark embedded data rate of the proposed method is doubled compared with the previous patchwork schemes [2, 22].

2) Using psychoacoustic model in wavelet packet domain can produce more precise masking thresholds. Incorporating these thresholds into our watermarking scheme objectively ensures that there is no audible quality degradation on audio signal after embedding watermarking.

3) To adaptively decide watermark strength according to masking thresholds and statistical properties of the host signal, the proposed algorithm can survive attacks, such as collusion and copy attack and Wiener filter attack, etc [3].

4) To embed watermark in tonal components makes the proposed signal sustainable to low bit rate compression and desynchronization attacks [6, 21].

   Desynchronization attacks include time-shifting, time-scaling, scrambling attack, etc. Such attacks prove to be very effective by SDMI [15, 21]. SDMI uses time scaling pitch-preserving attack or time scaling with 3% pitch increase attack to test the robustness of watermark algorithms against desynchronization attacks. There are two countermeasures to this kind of attacks: one is to estimate and undo the stretching, which is sometimes computation or time consuming; the other is to embed and detect watermark in a domain that is resilient to stretching/squeezing, as [11] does, which embeds watermark by changing the time-scale between two extrema (successive maximum and minimum pair) of the host signal. According to the same principle, we embed watermark in tonal components, which are resilient to time scaling of the host signal. Hence the proposed method is inherent immune to the time scaling attack.

Table 1. Detection error rate for different attacks (a. MPEG compression, b. Filtering, c Resampling, d. Time scaling (Preserving pitch), e. Time scaling (3% Pitch increase))

	Attacks	a	b	c	d	e
Error Detection Rate (10^{-5})	Blues1	28.33	28.09	27.76	27.82	28.10
	Blues2	28.30	28.08	27.81	27.88	28.05
	Classical1	28.25	28.12	27.89	28.16	27.85
	Classical2	28.17	27.97	27.78	28.04	27.88
	Country1	28.29	27.89	28.01	28.19	27.92
	Country2	28.21	27.94	28.08	28.11	28.02
	Folk1	28.27	27.86	27.83	27.98	28.09
	Folk2	28.22	27.92	27.69	27.87	27.96
	Pop1	28.16	28.06	27.92	28.06	27.45
	Pop2	28.19	28.11	27.94	28.10	27.68

In order to test the robustness of the proposed watermarking method to common signal manipulations as well as malicious attacks, a set of experiments were carried out. Ten wideband audio signals used for the performance evaluation of audio watermark algorithms in [7] were employed in this paper as host signals. All of the host signals are sampled at 44.1 kHz with 16-bit resolution. They represent five typical classes of music: blues, classical, country, folk, and pop. Each two-bit watermark was embedded 10 times. The size of the audio frame was 512 and the watermark dataset contained 100 (N) tonal components. The size of the subset *A* or *B* is 50(n). Watermarks were embedded at total 861 frames on end, approximately 10 seconds.

The common signal processing operations and malicious attacks include:

1) MPEG 1 Layer III audio compression: To test the robustness of the proposed scheme to low bit-rate coding, MPEG 1 layer III algorithm with compression rate of 96kbps was applied to the watermarked signals.

2) Filtering: a 15-tap low-pass filter with a cutoff frequency of 441Hz was employed.

3) Resampling: Watermarked signals at 44.1 kHz sampling rate were subsampled down to 22.05 kHz and interpolated back to 44.1 kHz.

4) Cropping: continuously embedded the same two-bit watermark message 10 times enables correct detection after a small amount of signal (less than 116ms) is cropped or cut.

5) Time scaling attacks: speeding up the watermarked signals by 5% while preserving pitch and speeding up by 5% with 3% pitch increase were tested.

The experimental results in Table 1 show that the proposed watermarking method can successfully survive various types of signal manipulations and attacks. The artifacts introduced by watermark insertion are inaudible; therefore the quality of the watermarked signal isn't subjectively degraded.

5. CONCLUSION

This paper presents a novel audio watermarking algorithm which takes advantage of both auditory and statistical properties to design the scheme and embed watermark. Since the new scheme incorporates psychoacoustic model analysis in time and wavelet packet domain, it can adapt the watermark strength to host signal while maintain transparency of the whole watermarking process.

The proposed watermarking scheme also adopts some measures to improve robustness against various signal processing operations and attacks, including adaptively embedding, using perceptual significant portions to embed watermark, etc.

False detection analysis and experimental results demonstrate the proposed watermarking scheme has good robustness and inaudibility; therefore it is a promising method to protect the copyright of the audio signals.

6. ACKNOWLEDGMENTS

The work described in this paper was supported by two grants, the National Science Foundation of China No. 60075002 and

National High-tech R&D Program (863 Program) of China No. 2003AA144080.

7. REFERENCES

[1] Ambikairajah E., Davis A. G., Wong W. T. K. Auditory masking and MPEG-1 audio compression. *IEEE J. Electronics & Communication Engineering*, vol. 9, no. 4, 1997, 165-175.

[2] Arnold M. Audio watermarking: Features, applications and algorithms. In *IEEE Int. Conf. Multimedia and Expo 2000*, vol. 2, 2000, 1013-1016.

[3] Arnold M. Attacks on digital audio watermarks and countermeasures," in *IEEE Int. Conf. WEB Delivering of Music*, 2003, 1-8.

[4] Bender W., Gruhl D., Morimoto N., Lu A. Techniques for data hiding. *IBM Syst. J.*, vol. 35, no. 3/4, 1996, 313-336.

[5] Black M., Zeytinoglu M. Computationally efficient wavelet packet coding of wide-band stereo audio signals. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1995, 3075-3078.

[6] Cox I., Kilian J., Leighton T., Shamoon T. Secure spread spectrum watermarking for multimedia. *IEEE Trans. Image Processing*, vol. 6, 1997, 1673-1687.

[7] Gordy J. D., Bruton L. T. Performance evaluation of digital audio watermarking algorithms. In *43rd IEEE Midwest Symposium on Circuits and Systems*, , vol.1, 2000, 456-459.

[8] ISO/IEC IS11172-3. Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s—Part 3: Audio. 1992.

[9] Kundur D., Hatzinakos D. Toward robust logo watermarking using multiresolution image fusion principles. *IEEE Trans. Multimedia*, vol.6, no.1, 2004, 185-194.

[10] Lemma A.N., Aprea J., Oomen W., Leon van de Kerkhof. A temporal domain audio watermarking technique. *IEEE Trans. Signal Processing*, vol. 51, no.4, 2003, 1088-1097.

[11] Mansour M. F., Tewfik A. H. Data embedding in audio suing time-scale modification *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 3, 2005, 432-440.

[12] Podilchuk, C.I., Delp, E.J. Digital watermarking algorithms and applications. *IEEE Signal Processing Magazine*, vol. 18, no.4, 2001, 33-46.

[13] Pan D. Y. A tutorial on MPEG/audio compression. *IEEE Multimedia*, vol. 2, no. 2, 1995, 60-74.

[14] Quan X. M., Zhang H. B. Audio watermarking for copyright protection based on psychoacoustic model and adaptive wavelet packet decomposition. In *Proceedings of 2004 international symposium on Intelligent Multimedia, Video and Speech Processing*, 2004, 282-285.

[15] Secure Digital Music Initiative: http://www.sdmi.org.

[16] Sinaga F., Gunawan T. S., Ambikairajah E. Wavelet packet based audio coding using temporal masking. In *ICICS-PCM 2003*, 2003, 1380-1383.

[17] Sinha D., Tewfik A. Low bit rate transparent audio compression using adapted wavelets. *IEEE Trans. Signal Processing*, vol.41, no. 12, 1993, 3464-3479.

[18] Srinivasan P., Jamieson L. H. High quality audio compression using and adaptive wavelet packet decomposition and psychoacoustic modeling. *IEEE Trans. Signal Processing*, vol. 46, 1998, 1085-1093.

[19] Tachibana R., Shimizu S., Kobayashi S., Nakamura T. An audio watermarking method using a two-dimensional pseudo-random array. *Signal Processing*, vol.82, 2002, 1455-1469.

[20] Tzeng J.N., Hwang W. L., Chern I. L. An asymmetric subspace watermarking method for copyright protection. *IEEE Trans. Signal Processing*, vol.53, no.2, 2005, 784-792.

[21] Wu M., Craver S. A., Felten E. W., Liu B. Analysis attacks on SDMI audio watermarks. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, 2001, 1369-1372.

[22] Yeo I. K., Kim H. J. Modified patchwork algorithm: A novel audio watermarking scheme. *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 4, 2003, 381-386.