14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A HYBRID PRE-WHITENING TECHNIQUE FOR DETECTION OF ADDITIVE SPREAD SPECTRUM WATERMARKS IN AUDIO SIGNALS

Krishna Kumar S.

Broadcast & Communications Group
Centre for Development of Advanced Computing
Trivandrum-695 033, INDIA.
E-mail: krishku@cdactvm.in

Thippur Sreenivas

Department of Electrical Communication Engineering
Indian Institute of Science
Bangalore - 560 012, INDIA.
E-mail: tvsree@ece.iisc.ernet.in

ABSTRACT

Pre-whitening techniques are employed in blind correlation detection of additive spread spectrum watermarks in audio signals to reduce the host signal interference. A direct deterministic whitening (DDW) scheme is derived in this paper from the frequency domain analysis of the time domain correlation process. Our experimental studies reveal that, the Savitzky-Golay Whitening (SGW), which is otherwise inferior to DDW technique, performs better when the audio signal is predominantly lowpass. The novelty of this paper lies in exploiting the complementary nature to the two whitening techniques to obtain a hybrid whitening (HbW) scheme. In the hybrid scheme the DDW and SGW techniques are selectively applied, based on short time spectral characteristics of the audio signal. The hybrid scheme extends the reliability of watermark detection to a wider range of audio signals.

1. INTRODUCTION

Additive spread spectrum watermarking is a very established technique in the field of audio watermarking. Typically a psychoacoustically shaped pseudo random (PR) sequence is added to the *host* audio signal (either directly in the time domain or in one of the transformed domains) and subsequently detected using linear correlator (matched filter) detection techniques. When blind detection is performed using a linear correlator, there exists the problem of the undesirable correlation between host signal and the watermark at the detector. Analogous to a communication channel, the host signal is traditionally visualized as an AWGN channel in which the watermark message is communicated. However, real audio signals do not have white noise properties as adjacent audio samples are highly correlated, particularly when short-time segments of audio signal are used for detection. It is known from the theory of statistical signal detection that the simple matched filter is suboptimal when the noise is correlated.

The well known *generalized matched filter* detectors [1] circumvent this problem by accounting for the correlated nature of the noise (here the audio signal). The generalized matched filter (GMF) is obtained when a likelihood ratio test (LRT) is applied to the detection of a known signal corrupted by a wide sense stationary *correlated* Gaussian random noise. The GMF can be visualized as a *pre-whitening* followed by a simple matched filter. In this case the whitening is performed in a statistical sense based on the noise covariance matrix. Kim [2] suggest a watermarking scheme based on this principle. It is also possible to perform the pre-whitening explicitly and then use a simple matched filter detector. Some of the techniques suggested in the literature exploits the white spectral characteristics of the linear prediction residual [3] or the Savitzky-Golay residual [4] to accomplish audio pre-whitening.

Haitsma et. al. [5] uses a technique, called *symmetric phase-only matched filter* (SPOMF), for correlation detection of additive watermarks. In this scheme, the test vector is *deterministically* whitened before computing the correlations. Since whitening tends to improve the detection, they argue, that one may take it to the extreme case in which the magnitude spectrum of the test vector is forced to unity. One of the contributions of this paper (Section 3) is the derivation of the deterministic whitening scheme from the *the frequency-domain view* of the time-domain correlation process. We believe that this treatment is more intuitively appealing, particularly for the audio signal processing community. The scheme, which we would denote as *direct deterministic whitening* (DDW) is, in principle, similar to the SPOMF technique.

Though deterministic whitening is expected to give better results compared to statistical whitening techniques, our experimental studies reveal that the spectral properties of the audio signal influence the effectiveness of the whitening techniques. Though, the DDW technique performance is better in general, it is observed that for a certain class of signals, where DDW technique performance is unreliable, the *Savitzky-Golay whitening* (SGW) technique [4] performs well. The second contribution of the paper (Section 4) is the design of a *hybrid whitening* (HbW) scheme, based on this complementary nature of the two whitening techniques. In the hybrid whitening scheme, the DDW and SGW techniques are selectively applied, based on short time spectral characteristics of the audio signal.

2. THE WATERMARKING SCHEME

In this discussion we confine our attention to *additive spread spectrum watermarking in the time domain*. It is known from the SS theory, that *pseudo random* (PR) sequences, generated from user specific keys, are good choices as watermarks. The the spectral and temporal distribution of the watermark energy is shaped such that it remains inaudible in the host signal. A correlator detector can be used for the detection of a specific PR sequence in any given audio clip.

The first author is affiliated also to the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore as a research scholar.

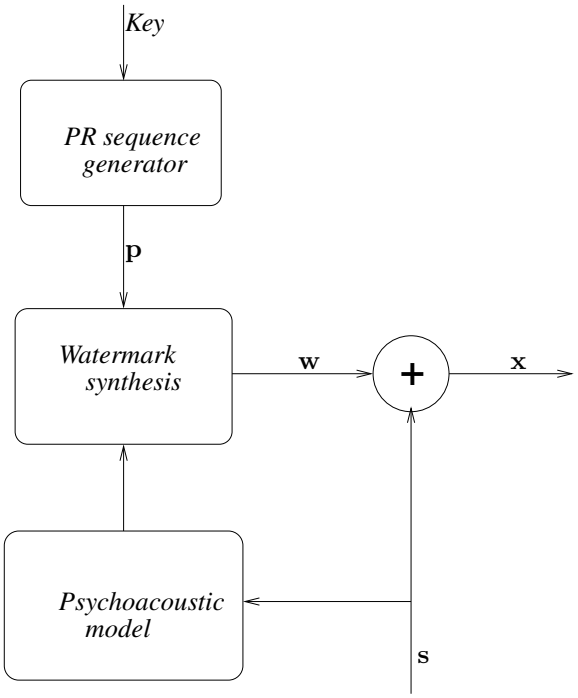

Figure 1: *Watermark Embedding Scheme*

2.1 Watermark Embedding

Given an N-length host vector $\mathbf{s} \equiv s[n]; 0 \leq n \leq N-1$, we design a watermark vector $\mathbf{w} \equiv w[n]; 0 \leq n \leq N-1$ and embed it additively with $\mathbf{s}$. The watermarked vector $\mathbf{x} \equiv x[n]; 0 \leq n \leq N-1$ may be expressed as

$$\mathbf{x} = \mathbf{s} + \mathbf{w}. \tag{1}$$

The watermark sequence $\mathbf{w}$ is computed from $\mathbf{s}$ and a *pseudo random* (PR) sequence $\mathbf{p}$, by a *shaping* process as

$$\mathbf{w} = \mathbf{\Psi}(\mathbf{p}, \mathbf{s}) \tag{2}$$

so as to keep $\mathbf{x}$ perceptually identical to $\mathbf{s}$. We denote this process by the term *watermark synthesis*. The watermark synthesis function $\mathbf{\Psi}(\mathbf{p}, \mathbf{s})$ is defined implicitly as

$$W[k] = M_S[k] e^{j\theta_P[k]}; 0 \leq k \leq N-1 \tag{3}$$

where $W[k]$ is the N-point DFT's of $w[n]$ and $M_S[k]$ is the psychoacoustically computed *masking profile* for the host segment $\mathbf{s}$. In this work we used the MPEG-1 psychoacoustic model-1 [6] to derive the masking profile. The scheme is depicted in Fig. 1.

2.2 Need for Random Phase Watermarks

In audio watermarking, the requirement of perceptual transparency forces the watermark power spectrum to follow the spectral masking profile of the host signal. Further, it is pointed out [7] that, a watermark is *energy efficient* and robust to *Wiener attack* if the watermark power spectrum is proportional to the host power spectrum. Under these conditions, the watermark power spectrum is decided by the host signal and the randomness of the watermark is contributed by its *phase spectrum*. In view of these constraints, we have to use *random phase sequences* as watermarks [8]. A real-valued random phase sequence $\mathbf{p} \equiv p[n]; 0 \leq n \leq N-1$ can be generated, whose phase spectrum $\theta_P[k]$ satisfying the following conditions:

$$\theta_P[k] \in \{0, \pi\}; \; k = 0, \frac{N}{2} \tag{4}$$

$$\theta_P[k] \in (-\pi, +\pi]; \; 1 \leq k \leq \frac{N}{2} - 1 \tag{5}$$

$$\theta_P[k] = -\theta_P[N-k]; \; 0 \leq k \leq N-1 \tag{6}$$

The time-domain sequence $p[n]$ is given by

$$p[n] = \mathscr{F}_N^{-1}\left(e^{j\theta_P[k]}\right); n, k \in \{0, 1, \ldots N-1\} \tag{7}$$

Here, $\mathscr{F}_N^{-1}$ denotes the N-point IDFT operator. Note that the magnitude spectrum of this sequence is white, which can be appropriately shaped by the watermark embedder.

2.3 Watermark Detection using Correlator-Detector

At the stage of watermark detection, let the received signal be $\mathbf{r}$, with or without a watermark embedded. The task of the watermark detector is to verify the presence of a particular watermark sequence $\mathbf{w}$, which is characterized by the specific random phase sequence $\mathbf{p}$. The vectors $\mathbf{r}$ and $\mathbf{p}$ reside in the vector space $\mathscr{R}^N$. A correlator-detector can be used to compute the correlation ρ_{rp} between $\mathbf{r}$ and $\mathbf{p}$ as

$$\rho_{rp} = \frac{1}{N} \sum_{n=0}^{N-1} r[n] p[n] = \frac{1}{N} \mathbf{r}^T \mathbf{p} \tag{8}$$

based on which, the following binary hypotheses testing is performed:

Hypothesis H1: Legitimate watermark present;

$$\mathbf{r} = \mathbf{x} = \mathbf{s} + \mathbf{w}. \tag{9}$$

Since $\mathbf{s}$ is a segment of the audio signal, which comes from a large class of signals, it can be considered as a random vector, and the test statistic ρ_{rp} can be modeled as a random variable with respect to $\mathbf{s}$. Let us denote the test statistic under this hypothesis as $(\rho_{rp}; H1)$ with the pdf $f(\rho_{rp}; H1)$.

Hypothesis H0: Legitimate watermark absent;

$$\mathbf{r} = \mathbf{y} = \mathbf{s} + \mathbf{v}, \tag{10}$$

where $\mathbf{v}$ is any N-length watermark vector (including the *zero vector*), such that $\mathbf{v} \neq \mathbf{w}$. Again, the test statistic under this hypothesis, denoted as $(\rho_{rp}; H0)$, is a random variable with a pdf $f(\rho_{rp}; H0)$. Let this be the *alternate* pdf.

For detection, the given instance of ρ_{rp} is compared with an optimum threshold γ, and one of the hypotheses is decided. In practice, the correlation ρ_{rp} is normalized with $\|\mathbf{r}\| = \sqrt{\sum_{n=0}^{N-1} |r[n]|^2}$ to counteract the variations in the received signal strength.

3. NEED FOR PRE-WHITENING

In this section, we derive the deterministic whitening scheme from the *the frequency-domain view* of the time-domain correlation process.

The correlation defined by Eqn. (8) may be expressed in the DFT-domain as

$$\rho_{rp} = \frac{1}{N}\sum_{k=0}^{N-1} |R[k]|\,|P[k]|\cos(\theta_R[k]-\theta_P[k]). \qquad (11)$$

where $\theta_R[k]$ and $\theta_P[k]$ are the phase spectra of $r[n]$ and $p[n]$ respectively, and $|R[k]|$ and $|P[k]|$ are the corresponding magnitude spectra.

An arbitrary signal $r[n]; 0 \le n \le N-1$, at the detector can be expressed as

$$r[n] = s[n] + u[n] \qquad (12)$$

where $r[n]$ is the host signal and $u[n]$ is an arbitrary watermark signal (legitimate, alternate or zero). The correlation between $r[n]$ and $p[n]$ can be expressed as:

$$\begin{aligned}\rho_{rp} &= \frac{1}{N}\sum_{n=0}^{N-1} s[n]p[n] + \frac{1}{N}\sum_{n=0}^{N-1} u[n]p[n] \\ &\triangleq \rho_{sp} + \rho_{up} \qquad (13)\end{aligned}$$

where ρ_{sp} is the *watermark-to-host correlation* (WHC) and ρ_{up} is the *watermark-to-watermark correlation* (WWC).

It is desirable that ρ_{sp} has as low a variance as possible, with respect to $\mathbf{s}$. This permits the detector to make a judgment based mainly on the value of ρ_{up}, which is the correlation between the watermark of interest and the watermark which is (possibly) embedded in the test signal.

By applying equation (11) on ρ_{sp}, and since $|P[k]| \triangleq 1$, we have

$$\rho_{sp} = \frac{1}{N}\sum_{k=0}^{N-1} |S[k]|\cos(\theta_S[k]-\theta_P[k]) \qquad (14)$$

It can be immediately seen that ρ_{sp} is the N-sample average of the random sequence $|S[k]|\cos(\theta_S[k]-\theta_P[k])$. It is possible (under certain conditions) to approximate the sample average of a random sequence with the corresponding ensemble average. *However*, this approximation is valid only in the limiting sense i.e. as $N \to \infty$, where N is the length of the sequence. For finite N, the sample-average itself can be considered as a random variable, whose expectation is the ensemble average and whose variance is *inversely related* to N. In general, the magnitude spectrum $|S[k]|$ of an audio signal, has a large variability and has several (let us say K) local spectral peaks. Fig. 2 shows the magnitude spectrum of a 23ms frame of a typical music signal. We have used the linear scale (as opposed to the conventional dB scale) to clearly bring out the peaky nature the magnitude spectrum. The relatively small number of these dominant components (i.e. $K \ll N$) determine the overall correlation, resulting in increased $\sigma^2_{\rho_{sp}}$, which leads to poor detection performance.

Let us consider a sequence $\tilde{p}[n]$ with a corresponding N-point DFT $\tilde{P}[k]$. Define $\tilde{P}[k]$ such that

$$\tilde{P}[k] \triangleq \frac{1}{|S[k]|}e^{j\theta_P[k]} \quad \forall k \qquad (15)$$

If the new sequence $\tilde{p}[n]$ is used in the correlation process, in place of the original PR sequence $p[n]$, the new watermark-to-host correlation $\rho_{s\tilde{p}}$ is obtained , using Eqn.(11), as

$$\rho_{s\tilde{p}} = \frac{1}{N}\sum_{k=0}^{N-1} |S[k]|\,\frac{1}{|S[k]|}\cos(\theta_S[k]-\theta_P[k])$$

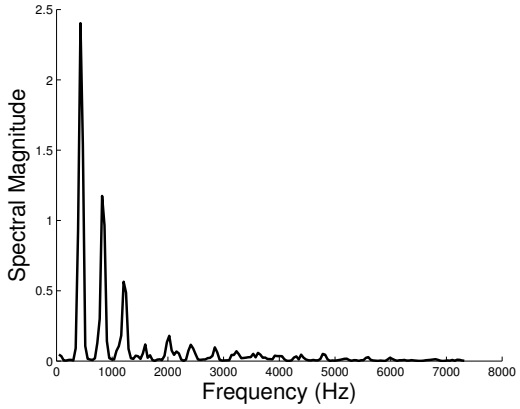

Figure 2: *Magnitude spectrum plot of a 23ms frame of a typical music signal, in linear magnitude scale.*

$$= \frac{1}{N}\sum_{k=0}^{N-1} \cos(\theta_S[k]-\theta_P[k]) \qquad (16)$$

It may be observed that the influence of the host magnitude spectrum is completely removed in this formulation. This is particularly useful in the context of random phase watermarks, since the watermark magnitude spectrum is dictated completely by the host signal characteristics and contains no information about the underlying PR sequence.

In the blind detection scenario, (15) cannot be implemented exactly since the host magnitude spectrum $|S[k]|$ is not available at the detector. However, $|S[k]|$ may be substituted by $|R[k]|$, which is a close approximation to $|S[k]|$. Using this approximation, the correlation between $r[n]$ and the new sequence $\tilde{p}[n]$ can be expressed as:

$$\rho_{r\tilde{p}} = \frac{1}{N}\sum_{k=0}^{N-1} \cos(\theta_R[k]-\theta_P[k]). \qquad (17)$$

The above approach to reduce the variance of the test statistic can be interpreted as a whitening of the (possibly) watermarked signal $r[n]$. Let $\tilde{r}[n]$ be the whitened version of $r[n]$. Then its DFT, by definition, is given by

$$\tilde{R}[k] \triangleq e^{j\theta_R[k]}; 0 \le k \le N-1. \qquad (18)$$

Now, the correlation between $\tilde{r}[n]$ and $p[n]$ is

$$\rho_{\tilde{r}p} = \frac{1}{N}\sum_{n=0}^{N-1} \tilde{r}[n]p[n] = \frac{1}{N}\sum_{k=0}^{N-1} \cos(\theta_R[k]-\theta_P[k]), \qquad (19)$$

which is identical to $\rho_{r\tilde{p}}$ given by Eqn. (17). Thus, the *deterministic whitening* process, equalize the magnitude spectrum of the audio signal, so that $K \approx N$, thus keeping $\sigma^2_{\rho_{sp}}$ to a desirable level. It may also be noted that $\rho_{\tilde{r}p}$, being a function of phase values only, is *independent of the signal strength*.

The direct realization of Eqn.(18) is a feasible whitening technique in the watermark detection scheme. Since the magnitude spectrum of $\tilde{r}[n]$ is *deterministically* forced to unity, we denote this technique by the term *direct deterministic whitening* (DDW). A block schematic of watermark detection scheme, using whitening and correlation detection, is shown in Figure 3.

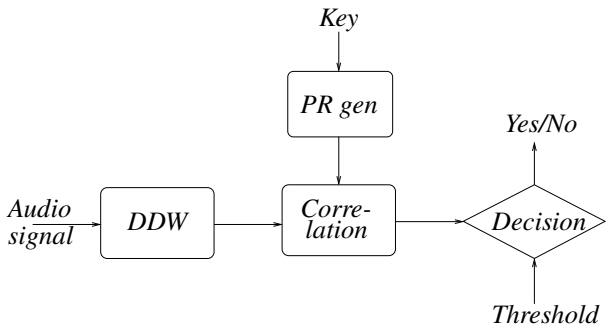

Figure 3: *Watermark Detection with Pre-whitening*

4. HYBRID PRE-WHITENING SCHEME

We have seen that a number of techniques are available for whitening the audio signal in the statistical sense. Of particular interest to us is a technique suggested by Cvejic and Seppanen [4], which we denote as the *Savitzky-Golay whitening* (SGW) technique. The Savitzky-Golay filters [9] are a class of smoothing filters based on polynomial fitting of the (audio) data. It is observed that the residual of this smoothing operation tends to be white Gaussian. Since the DDW technique whitens the signal to the extreme, it is expected to give better results compared to statistical whitening techniques. Though this is true in general, our experimental studies reveal that, for a certain class of signals, the *Savitzky-Golay Whitening* (SGW) technique outperforms DDW.

To illustrate this feature, we use the audio clip 'bass1.wav' which is sampled at 44100 Hz and represented using16-bit linear PCM. The clip is watermarked using the embedding algorithm described in section 2.1. The watermarked as well as the un-watermarked versions of the signal are subject to watermark detection using the scheme depicted in Figure 3. To achieve reliable detection, the basic scheme is modified as follows. The correlations were calculated over segments of length $N = 4096$ and the results are averaged over the first $B = 16$ segments having sufficient level. To counteract any mis-synchronization of the received signal, the correlations are calculated as circular cross correlations over N circular shifts. The averaging was also done for each shift. Further, the individual correlations, i.e. before averaging, were normalized by the norm of the whitened signal. (This will not make any difference in DDW while it helps the SGW scheme, in which case the whitening may not be perfect.) The highest value of the averaged correlation coefficient (*maximum averaged correlation coefficient*) is considered as test statistic $\rho_{\tilde{r}p}$. To make the system robust to lowpass filtering and AWGN attacks, only the frequencies below 8kHz were used for watermarking.

Figure 4 shows the time-domain plot of the signal (top) as well as its spectrogram (bottom), computed over 23ms long frames with 50% overlap. The frequency range is limited to 8kHz, since the watermark detection is confined to this region of the spectrum. It can be immediately observed that the first half of the signal is lowpass in nature and is confined to a small part of the audio spectrum.

The first (top) and second (middle) plots in Figure 5 shows the correlation values obtained using the SGW and the DDW techniques respectively. The Savitzky-Golay filters use 4^{th} order polynomials and a frame size of 21 as suggested

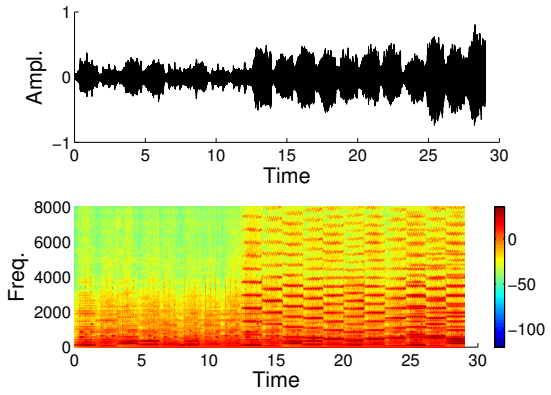

Figure 4: *Time-Frequency Representation of audio signal 'bass1.wav'. Signal plot (top) and Spectrogram (bottom).*

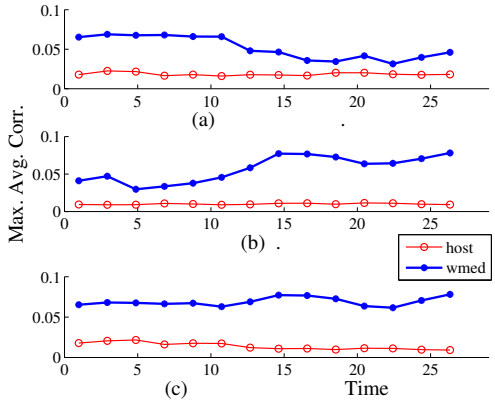

Figure 5: *Correlations plot for the signal 'bass1.wav' - with (a) SGW (b) DDW and (c) HbW. The thick (blue) line is for watermarked signal and the thin (red) line is for host signal.*

in [4]. Clearly, the DDW outperforms Savitzky-Golay technique in the second half of the signal. However, in the first part, *where the audio signal is predominantly lowpass*, the DDW performance is below par, but the SGW performance is reliable. *Motivated by this complementary behavior of the two whitening techniques*, we propose a *Hybrid Whitening* (HbW). For every N-length segment of the audio signal, a *spectral weight ratio* ζ is computed and a suitable threshold ν is chosen, where ζ is defined as

$$\zeta = \frac{\sum_{k=L}^{N-1} |R[k]|^2}{\sum_{k=0}^{L-1} |R[k]|^2}; \quad 0 < L < N-1. \tag{20}$$

For any given segment, if $\zeta < \nu$, that segment is pre-whitened using the SGW technique. If $\zeta \geq \nu$, the segment is pre-whitened using the DDW technique. The parameters L and ν are to be appropriately chosen to maximize the advantage. The third (bottom) plot in Figure 5 shows the correlation values obtained using the HbW technique. It may be readily observed that the performance of the hybrid scheme is better than those of either of the individual techniques.

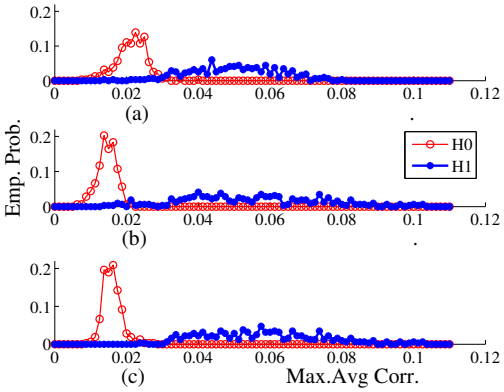

Figure 6: *Empirical distributions of maximum averaged correlations - with (a) SGW (b) DDW and (c) HbW. The thick (blue) line is* $(f_{\rho_{\tilde{r}p}};H1)$ *and the thin (red) line is for* $(f_{\rho_{\tilde{r}p}};H0)$.

5. EXPERIMENTAL STUDY

Having obtained the technique, its utility has to be verified experimentally. Towards this, an ensemble of 60 polyphonic audio signals were collected. Each was of duration 24 seconds, and in the 44100 Hz, 16 bits/sample, single channel (mono) format. The signals were watermarked and from this ensemble, 316 audio clips of 4 seconds duration were extracted, ignoring the low energy portions and silence regions. Each clip was then subject to the watermark detection method used in the case of the example clip (Section 4). The test statistic $\rho_{\tilde{r}p}$ is evaluated separately for the SGW, DDW and the HbW schemes, for both legitimately watermarked signals (hypothesis $H1$) as well as un-watermarked signals (hypothesis $H0$). The distributions of the test statistic are given in Figure 6.

The *Fisher's Discriminant Ratio* F can be used for quantifying the detection performances. This ratio is defined as

$$F = \frac{[(\mu;H1)-(\mu;H0)]^2}{(\sigma^2;H0)+(\sigma^2;H1)} \qquad (21)$$

where $H0$ and $H1$ are the two hypotheses, μ denotes the mean and σ denotes the standard deviation under each hypothesis. In this study, we use the sample mean and sample standard deviation obtained form the empirical data, to compute F. Table 1 shows the sample means and standard deviations for each technique under each hypothesis. The right most column lists the Fisher's ratio F which characterizes the detection performance. The results show that, the SGW and DDW techniques, when applied individually give nearly identical performance, while the hybrid scheme provides a clear improvement .

6. CONCLUSION

In this paper, we have derived the direct deterministic whitening (DDW) scheme to improve the detection performance of the correlation detection of additive random-phase watermarks in audio signals. The derivation was motivated by the frequency domain view of the correlation process and we believe this more intuitively appealing compared to the statistical treatment. Further, it was observed that the performance of of the Savitzky-Golay whitening (SGW) method is *selectively* better than that of the deterministic whitening method. The Savitzky-Golay whitening (SGW) is found to outperform if the audio signal is predominantly lowpass. Importantly, the DDW and SGW performances are found to be complementary and a *hybrid whitening scheme* is derived taking advantage of this property. The hybrid scheme outperforms the individual techniques, *particularly when applied over a wide variety of audio signals*. Thus the hybrid scheme extends the reliability of watermark detection to a wider range of audio signals.

Table 1: Comparison of Detection Performances

Scheme	$(\mu;H1)$	$(\sigma;H1)$	$(\mu;H0)$	$(\sigma;H0)$	F
SGW	0.0215	0.0499	0.0044	0.0121	4.91
DDW	0.0145	0.0546	0.0026	0.0185	4.64
HbW	0.0159	0.0578	0.0026	0.0160	6.67

REFERENCES

[1] S. M. Kay, "Generalized Matched Filters" in *Fundamentals of Statistical Signal Processing, Vol. 2 - Detection Theory*, Prentice Hall PTR, 1993, Chap. 4, pp.105 – 112.

[2] H. Kim, "Stochastic Model based Audio Watermark and Whitening Filter for Improved Detection", *Proc. IEEE Intl. Conf. Acoust. Speech and Signal Proc.*, Apr 2000, pp 1971 – 1974.

[3] J. W. Seok and J. W. Hong, "Audio Watermarking for Copyright Protection of Digital Audio Data", *IEE Electronic Letters*, Vol. 37, No.1, 4th Jan 2001, pp. 60-61.

[4] N. Cvejic and T. Seppanen, "Audio Prewhitening based on Polynomial Filtering for Optimal Watermark Detection", *Proc. Eur. Signal Proc. Conf.*, Sept 2002.

[5] J. Haitsma, M. van der Veen, T. Kalker and F. Bruekers, "Audio Watermarking for Monitoring and Copy Protection", *Proc. ACM Multimedia Workshop-2000*, pp 119 – 122.

[6] International standard: *ISO/IEC-11172-3, Information Technology* — "Coding of Moving Pictures and Associated Audio for Digital Storage Media up to about 1.5 Mbit/s — Part 3: Audio", 1993.

[7] J. K. Su and B. Girod, "Power-Spectrum Condition for Energy Efficient Watermarking", *IEEE Trans. Multimedia*, Vol. 4, No. 4, December 2002, pp. 551-560.

[8] L. Gang, A. N. Akansu and M. Ramkumar, "Security and Synchronization in Watermark Sequence", *Proc. IEEE Intl. Conf. Acoust. Speech and Signal Proc.*, Apr 2002, pp IV-3736 – IV-3739.

[9] S. J. Orfanidis, "Savitzky-Golay Smoothing Filters" in *Inrtoduction to Signal Processing*, Prentice Hall, 1996, Chap. 8, pp. 434 – 462.