APPLICATION OF SIDE-INFORMED EMBEDDING AND POLYNOMIAL DETECTION TO AUDIO WATERMARKING

M. Mullarkey[†]*, N. J. Hurley*[†]*, G. C. M. Silvestre*[†] *and T. Furon*[‡]

[†]University College Dublin, Belfield, Dublin 4 – Ireland
[‡]IRISA/INRIA, Campus de Beaulieu, Rennes – France

ABSTRACT

Spread sprectrum watermarking proceeds by extracting a feature vector from the cover contents and embedding a pseudo-random watermark signal in that feature vector. To detect the presence of the watermark, a correlation of the feature vector with the pseudo-random signal is performed and the result compared to a threshold. This correlation detection function is a first-order function of the feature vector components. In recent work, we have proposed that higher-order polynomial detection functions, combined with a side-informed watermark embedding strategy, can be used to increase the efficiency of the watermarking system. We have demonstrated this through a statistical analysis. In this paper, we apply our new family of detection functions to the watermarking of real audio signals. The schemes are tested on a database of over 300 different audio signals and a robustness analysis is performed on the experimental results.

1. INTRODUCTION

Watermarking hides information in a host signal, which can be extracted by an authorised detector without altering the perceptible quality of the host. Audio watermarks should be inaudible, robust to signal processing, such as compression, filtering or resampling and secure to malicious attempts to remove them. Digital rights management systems require *blind* watermark detection i.e. detection without recourse to the original host signal. Watermarking may be modelled as a communication problem, in which the host signal constitutes the channel for transmission of the watermark data. The watermark is modulated on a feature vector extracted from the host signal in the time domain or in some transform domain. The most popular form of blind watermarking, discrete sequence spread-spectrum watermarking (DSSS), embeds a pseudo-random sequence (chosen independently of the host) which is detected by application of a correlation detector. From a security point-of-view, one weakness of DSSS is that it is symmetric, in the sense that the watermark signal must be available to both the embedder and detector. An averaging attack can extract the watermark given a set of contents (of $O(N)$ size where N is the spreading length) marked with the same watermark signal.

For this reason, interest has been generated in the development of *asymmetric* schemes which do not require the watermark signal to be available to the detector. For example, [1, 2] propose second-order schemes in which the detector calculates a quadratic form on the extracted feature vector. These are more secure, since an attacker must estimate the quadratic form to remove the watermark which requires $O(N^2)$ attacks. This approach is generalised in [3], where an n^{th}-order detection function is proposed. In recent work [4, 5], we have demonstrated theoretically that by choosing a particular side-informed watermark embedding strategy, such n^{th}-order polynomial detection functions can also provide greater robustness, since they result in more powerful detection tests. Hence, our n^{th}-order schemes yield better performance for a given watermark embedding strength, than DSSS. Equivalently, for a given performance requirement (in terms of detection rate, probability of good detection and probability of false alarm), a smaller embedding strength is required for our n^{th}-order schemes than for DSSS and this in turn leads to greater security.

In this paper, we investigate the application of our n^{th}-order schemes to the watermarking of real audio signals. In particular, we examine some practical means by which feature vectors with desirable statistical properties can be extracted from audio, so as to yield the maximum detection power. Furthermore, we determine the value of n that yields the best performance, when a psychoacoustic model is used to determine the embedding strength. Finally, we test the performance of the schemes in the presence of AGWN attacks.

2. WATERMARK EMBEDDING AND DETECTION

The first stage of watermark embedding is to extract an N-dimensional feature vector $\mathbf{r}$ from the original cover data, $\mathbf{X}$, using an extraction function $e(\mathbf{X}, k) = \mathbf{r}$ which typically depends on a secret key k. The extraction process is invertible in the sense that there is an associated embedding process $m(\mathbf{X}, \mathbf{r})$ such that $m(\mathbf{X}, e(\mathbf{X}, k)) = \mathbf{X}$.

Given an embedding strength, g, an N-dimensional watermark vector $\mathbf{w}$ is mixed with the feature vector $\mathbf{r}$. We

0-7803-7663-3/03/$17.00 ©2003 IEEE

assume an additive mixing function, $F(\mathbf{r}, g\mathbf{w}) = \mathbf{r} + g\mathbf{w}$ where the watermark $\mathbf{w}$ is normalised to unit power. The watermarked content, $\mathbf{X}_w$ is then obtained via the embedding process: $\mathbf{X}_w = m(\mathbf{X}, F(\mathbf{r}, g\mathbf{w}))$.

Detection of the watermark from a received content $\tilde{\mathbf{X}}$, proceeds by extracting the feature vector $\mathbf{r} = e(\tilde{\mathbf{X}})$. We consider detectors that calculate a real-valued detection function $d(.)$ of the feature vector. The watermark is determined to be present if $d(\mathbf{r}) > thr$. for some threshold value thr. In discrete sequence spread spectrum (DSSS) watermarking, the detection function, $d(\mathbf{r}) = \mathbf{w}.\mathbf{r}$ where $\mathbf{w}$ is a pseudo-random signal.

2.1. Audio Extraction Functions

In our experiments, we analyse two different extraction functions, defined as follows:

2.1.1. Extraction Method 1

A fourier transform of length 1024 is applied to the time-domain audio samples. The frequency values are mapped into 25 *critical bands*, corresponding to frequency bands with similar auditory and masking properties. One component of $\mathbf{r}$ is extracted from each critical band CB and calculated as,

$$r_{\text{CB}} = \frac{1}{M} \sum_{i \in \text{CB}} \log_{10} |f_i| \,, \tag{1}$$

where M is the number of frequencies in the critical band. A psychoacoustic model can be used to determine the maximum distortion allowable on each critical band. Hence, the watermark strength can be varied from component to component, so as to maximise the strength of the watermark while remaining inaudible.

The watermark gw_{CB} is embedded back into the signal by updating the magnitudes of the frequency components,

$$|\hat{f}_i| = 10^{gw_{\text{CB}}} |f_i| \tag{2}$$

where $\hat{f}_i$ is the watermarked frequency component. An advantage of this method is that, since the watermark is duplicated on all frequencies in the critical band, there is some in-built robustness to desynchonisation by frequency cropping. With this method, to obtain a detection rate of one detection per second on CD-quality audio files, the watermark must be spread over a feature vector of length 1075.

2.1.2. Extraction Method 2

In this method, in order to generate a vector which is normally distributed, each component of the extracted vector is calculated as a sum of M randomly chosen frequency values. In particular, a set of k mid-band frequency values is chosen from each frame of a set of MN/k contiguous frames, giving a total of NM frequency values. These values are randomly partitioned into N sets of M frequency values and each component is generated as the sum of the log of the frequency magnitudes in each partition. In order to get good normality, a value of $M = 20$ was chosen.

In both cases, the components of the feature vector are permuted before the watermark is added, using a secret permutation. This permutation must be known to the detector.

3. DETECTION THEORY

Let $G = g^2/\sigma_r^2$ be the watermark to signal power ratio. In general, the output of the detection function is dependent on the embedding strength, G. Given $d(\mathbf{r})$, the goal is to distinguish between the null hypothesis H_0 that no watermark is present and the alternative H_1 that a watermark is present. The power function, $P_p(G) = P\{d(G) > thr|H_1\}$, gives the probability of correctly detecting the watermark given that the data is watermarked with embedding strength G and is in general an increasing function of G.

Let Q_0 be the centred and normalised cdf of the detection function d under H_0 and let Q_1 be the centred and normalised cdf of d under H_1. Let μ_0, σ_0 (resp. μ_1, σ_1) be the mean and standard deviation of the detection function under H_0 (resp. H_1). For a given probability of false alarm P_{fa}, the probability of good detection is given by

$$P_p\{d > thr|H_1\} = 1 - Q_1\left(\frac{\sigma_0}{\sigma_1} Q_0^{-1}(1 - P_{fa}) - e\right) \tag{3}$$

where the *efficiency* e is defined as $e = (\mu_1 - \mu_0)/\sigma_1$.

As Q_1 is an increasing function, increasing the efficiency of the detection results in increasing the power. The efficiency of the DSSS scheme for a watermark spread over N components is $\sqrt{NG}$.

4. SIDE-INFORMED EMBEDDING

The use of side-information has been recognised as a way to increase robustness of watermarking schemes [6, 7]. We have proposed a side-informed scheme, in which the watermark depends on the detection function. Given a detection function, we have shown in [4] that the expected output of the detector is maximised (to first order in $\sqrt{G}$) by choosing $\mathbf{w} \propto \nabla d(\mathbf{r})$. Moreover, in [4, 5], two families of polynomial detection functions of degree n were proposed.

$$\text{JANIS} \quad : \quad d(\mathbf{r}) = \sum_{i=1}^{N/n} \prod_{j=1}^{n} r_{i_j} \tag{4}$$

$$\text{POWER-n} \quad : \quad d(\mathbf{r}) = \sum_{i=1}^{N/2} r_i^{n-1} r_j \tag{5}$$

where j for POWER-n is a randomly chosen index, which is matched with each index i. To first order, it was shown that the efficiency of a JANIS detector, for normally distributed components of $\mathbf{r}$, is $\sqrt{nNG}$. The efficiency of a POWER-n detector also depends on the distribution of the components r_i. In particular, if the components of $\mathbf{r}$ are uniformly distributed, then POWER-n outperforms JANIS, with a first order efficiency of $n\sqrt{NG/6}$. Such large efficiency values can be traded-off against lower probability of false alarm, increased robustness to noise and/or lower embedding strengths.

5. FILTERING

From the above analysis, the performance of the different detection functions on a real audio signal depends on the distribution of the feature vector components. Before running the detection, the vector components can be filtered through a non-linear mapping, $h : r_i \rightarrow h(r_i)$, such as proposed in [8]. In particular, we propose $h(r) = (r - Q_\Delta(r))\text{sign}(\sin(2\pi r_i/\Delta))$ where Q_Δ is a quantisation with step-size Δ. If Δ is sufficiently small, then the resulting filtered components are approximately uniformly distributed. However, if Δ is too small, then the watermark itself is filtered out. Hence, we must choose $\Delta \gg g$. Filtering also removes a large amount of the variance of the original signal, so that the effective watermark to signal ratio is increased, leading to increased efficiencies in the detector. Probability distribution functions (pdf) of the feature vector components, with and without filtering are shown in Fig. 1.

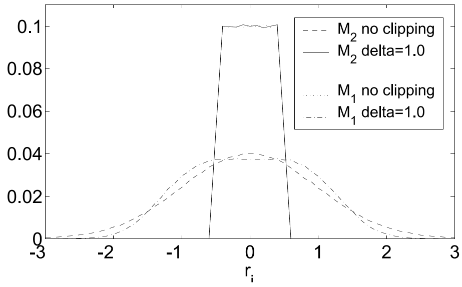

Figure 1: Distribution functions of the extracted feature vectors for the two extraction methods

6. EXPERIMENTAL RESULTS

The code is tested on a database of 350 30s audio clips, taken from a range of music types. One watermark symbol was spread over a 1s time period, using extraction methods 1 and 2 as described above.

6.1. The Effect of Filtering

The expressions for the efficiency given in Section 4 are first order in $\sqrt{G}$. When higher order terms of G are taken into account, the efficiency decreases for large n. The turning point (that is, the value of n at which the efficiency is maximised) depends on the value of G; the larger G is, the sooner the efficiency starts to decrease. Hence, higher order detection functions are useful when the power of the watermark is very low.

Filtering removes a significant portion of the variance of the signal and hence raises the effective watermark to signal ratio, G. Hence, if Δ is very small, the best detector will occur at $n = 2$. Small values of Δ will also ensure that the feature vector is uniformly distributed, which, from our theoretical results, indicates that a POWER-n detector is most effective. When Δ is larger, the feature vector becomes less uniform, which impacts on the performance of POWER-n in comparison to JANIS. Hence, there are a number of trade-offs to be made. A Δ must be chosen that yields good robustness to noise, then for this Δ, a detector must be chosen and its order set. In our experiments, we have chosen $\Delta = 1.0$, which results in increasing the effective watermark to signal ratio by about 10 dB.

6.2. Psychoacoustic Model

A psychoacoustic model is used to determine the masking properties of each critical band in the signal, and outputs a signal to mask ratio (SMR) for each critical band. By fixing the mask to noise ratio (MNR), a maximum power for the watermark is determined for each critical band and the frequencies within the band are marked with maximum power. Our experiments indicate that the watermark is inaudible when MNR=-13 dB. With G set in this way, the probability of good detection (P_p) is calculated for the JANIS and POWER-n detectors. For extraction method 1, it was be observed that the best value of n for the POWER-n detector is $n = 4$, and for JANIS is $n = 3$. JANIS out-performs POWER-n in this case. However, if the watermark is very weak, then higher orders of POWER-n outperform JANIS.

6.3. AWGN

The two extraction methods (labelled M1 and M2) are tested against AWGN attacks, with the noise to signal ratio (NSR) set in the range -40 dB to 0 dB and the watermark embedding strength set at G=-26 dB (see Fig. 2 and Fig. 3). Note that, with clipping, performance is very good at low noise, but falls sharply around -10 dB, when the noise is as strong as the filtered signal. Because the watermark to filtered signal ratio is so strong, POWER-n is no better than JANIS, even though the signal is uniform. Comparing Fig. 2 and Fig. 3 illustrates the importance of the distribution function of the extracted vectors. The second extraction method performs much better than the first, because vectors extracted using method 1 are not normally distributed. As expected from theory, POWER-n performs much more poorly than JANIS on the normally distributed vectors.

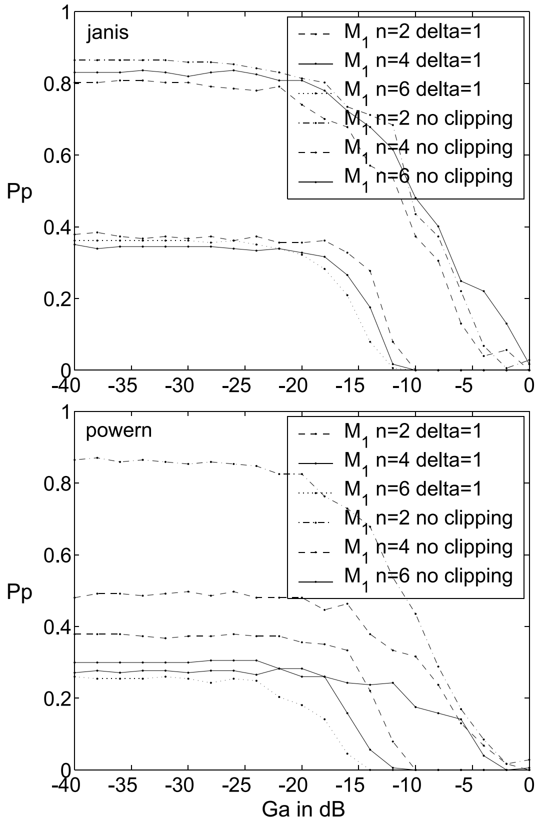

Figure 2: Extraction Method 1: Robustness to AWGN, P_{fa}=10^{-4}, G=-26 dB

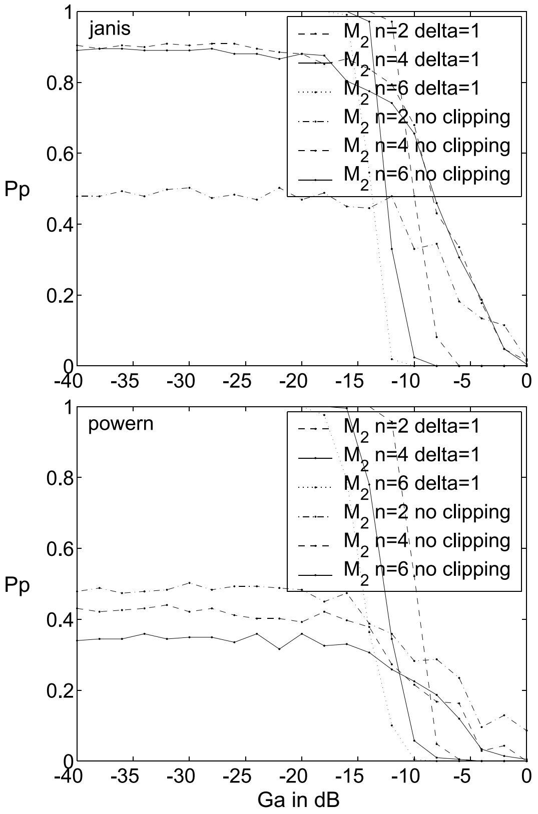

Figure 3: Extraction Method 2: Robustness to AWGN, P_{fa}=10^{-4}, G=-26 dB

7. CONCLUSION

We have presented results of a practical application of n^{th}-order side-informed watermarking to the watermarking of real audio signals. Using such schemes in practise requires that consideration be given to the distribution of the extracted feature vector, since this distribution determines the best detector to use. Although, theoretically, it was shown that POWER-n detectors can out-perform JANIS detectors on uniformly distributed vectors, the filtering method applied here to achieved uniformity, also boosts the power of the watermark in relation to the filtered signal. Hence, the watermark strength is too high to achieve good performance using high orders. High-order detection functions become most useful when a very low embedding strength is required.

8. REFERENCES

[1] J.J. Eggers, J.K. Su, and B.Girod, “Asymmetric watermarking schemes,” in *Tagungsband des GI Workshops Sicherheit in Mediendaten*, Berlin, Germany, Sept. 2000, Springer Reihe: Informatik Aktuell.

[2] T. Furon and P. Duhamel, “An asymmetric public detection watermarking technique,” in *Workshop on information hiding*, Dresden, Genmany, Oct. 2000.

[3] N.J. Hurley and G.C.M. Silvestre, “Nth order audio watermarking,” in *Proceedings of SPIE 2002*, January 2002.

[4] T.Furon, G. Silvestre, and N. Hurley, “Janis: Just another n-order side-informed scheme,” in *Proceedings of ICIP 2002*, October 2002.

[5] N.J. Hurley, G.C.M. Silvestre, and T. Furon, “Side-informed watermarking using nth-order polynomial detectors,” in *Proceedings of European Association for Signal Processing (EUSIPCO) XI European Signal Processing Conference 2002*, August 2002.

[6] I.J. Cox, M.L. Miller, and A.L. McKellips, “Watermarking as communication with side information.,” *Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information*, vol. 87, no. 7, pp. 1127–1141, July 1999.

[7] J.J. Eggers, J.K. Su, and B.Girod, “A blind watermarking scheme based on structured codebooks,” in *Proc. IEE Colloquium on Secure Images and Image Authentication*, Savoy Place London, April 2000, pp. 41–46.

[8] M. Ramkumar, *Data Hiding in Multimedia : Theory and Applications*, Ph.D. thesis, NJCMR, Princeton University, 2000.