A SYNCHRONISATION METHOD FOR INFORMED SPREAD-SPECTRUM AUDIO WATERMARKING

P.-Y. Fulchiron, F.B. O'Donovan, G.C.M Silvestre and N.J.Hurley

University College Dublin, Belfield, Dublin 4, Ireland

ABSTRACT

Under perfect synchronisation conditions, watermarking schemes employing asymmetric spread-spectrum techniques are suitable for copy-protection of audio signals. This paper proposes to combine the use of a robust psychoacoustic projection for the extraction of a watermark feature vector along with non-linear detection functions optimised with side-information. The new proposed scheme benefits from an increased level of security through the use of asymmetric detectors. We apply this scheme to real audio signals and experimental results show an increased robustness to desynchronisation attacks such as random cropping.

1. INTRODUCTION

The widespread adoption of broadband networks such as the Internet has rendered it economically feasible to distribute high-quality multimedia documents by electronic means without loss of quality. In this context, piracy and illicit dissemination has increased dramatically, threatening ownership and authors' rights. Digital watermarking is being considered as a promising technique in copy-control systems complementing cryptographic tools in the role of 'last-line of defense'. Most schemes proposed to date employ spread-spectrum techniques that can often be defeated by applying such malicious attacks as described in [1]. One very efficient, yet relatively simple attack consists of desynchronising the watermark detector by performing some random cropping, linear shifting or time warping imperceptible to the Human Auditory System (HAS).

In this paper, we propose the use of a robust psychoacoustic projection for the extraction of a feature vector as a means of synchronisation, along with an asymmetric detection scheme based on n^{th}-order polynomial forms using a side-informed method that has been shown to exhibit a higher level of security and better performance [2–4]. Computer simulations carried out over a large set of real audio data are used to evaluate the performance of the proposed scheme.

2. SPREAD-SPECTRUM WATERMARKING

Spread-spectrum watermarking (SSW) is commonly used to achieve blind detection. In most symmetric schemes, a watermark $\mathbf{w}$ is mixed with the original signal to produce the watermarked signal. Detection relies on applying an hypothesis test to the result of a correlation of the received signal with $\mathbf{w}$. Recently, increased robustness to collusion attacks was obtained with a number of asymmetric schemes of quadratic form as reported in [5]. However, security now relies on a secret permutation and, therefore, such schemes are more vulnerable to desynchronisation attacks. In [6], side-information was used to inform the watermark resulting in a large performance improvement.

2.1. Theoretical Framework

Given the original content $\mathbf{X_0}$, an N-dimensional feature vector $\mathbf{r_0}$ is extracted using an extraction function: $e(\mathbf{X_0}, k) = \mathbf{r_0}$, where k is a secret key. The extraction process is invertible in the sense that there is an associated embedding process $m(\mathbf{X}, \mathbf{r})$ such that $m(\mathbf{X}, e(\mathbf{X}, k)) = \mathbf{X}$. A central random watermark signal $\mathbf{w}$, normalised to unit power, is then modulated on $\mathbf{r}$ using a mixing function which can be assumed, without loss of generality, to be additive: $F(\mathbf{r_0}, g\mathbf{w}) = \mathbf{r_0} + g\mathbf{w}$. The resulting watermarked feature vector $\mathbf{r_w}$ is embedded back into the original content through the embedding process $m(.)$. In this section and for simplicity, the embedding strength is considered to be given by the constant scalar g.

Detection of the watermark from a received content proceeds by extracting an unknown feature vector $\mathbf{r}$ using extraction function $e(.)$, and testing the output of the detection function $d(\mathbf{r})$ under the null hypothesis, H_0, that no watermark is present, and the alternative, H_1, that a watermark is present. Generally, a threshold, thr, is set for a given probability of false alarm, $P_{fa} = P\{d(\mathbf{r}) > thr|H_0\}$, and the power of the test, $P_p = P\{d(\mathbf{r}) > thr|H_1\}$, is measured. Finally, the efficiency (also called deflection coefficient) for the scheme is defined as:

$$e_{d(\mathbf{r})} = \frac{E\{d(\mathbf{r})|H_1\} - E\{d(\mathbf{r})|H_0\}}{\sqrt{V\{d(\mathbf{r})|H_1\}}}$$

where $V\{.\}$ and $E\{.\}$ denote the variance and the expectation.

2.2. $\mathbf{n^{th}}$-Order Side-Informed Watermarking

In [3, 4], a renewed approach to watermarking was proposed by exploiting side-information at the embedder to maximise the detection power, and by using non-linear polynomial functions of n^{th}-order form for $d(\mathbf{r})$. Indeed, to first order and for small embedding strengths, the detection function can be written as:

$$d(\mathbf{r_0} + g\mathbf{w}) \simeq d(\mathbf{r_0}) + g\mathbf{w}^T \nabla d(\mathbf{r_0})$$

Side-information is then used to maximise the detection power by choosing the watermark $\mathbf{w} \propto \nabla d(\mathbf{r})$. Assuming that the detection function $d(\mathbf{r})$ is chosen such that $E\{d(\mathbf{r})|H_0\} = 0$, then the expected value of the detector under hypothesis H_1, that the watermark is present, is given to the first order by

$$E\{d(\mathbf{r})|H_1\} = g \sum_{i=1}^{N} \sqrt{E\left\{\left(\frac{\partial d(\mathbf{r})}{\partial r_i}\right)^2\right\}}.$$

In [3, 4], two families of detectors were proposed, namely JANIS and POWER-n. These have the best efficiency reported to date and are based on the following two polynomial functions of degree n:

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

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

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 $\sqrt{nNG}$, a factor of $\sqrt{n}$ better than SSW, where $G = \sigma^2_{\mathbf{r_0}}/g^2$. 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.

2.3. Feature Vector Choice

As we described in the previous sections, the extraction function $e(.)$ provides the vector $\mathbf{r_0}$ on which the watermark signal is embedded. The choice of the feature vector $\mathbf{r_0}$ is of great importance for the watermarking scheme as robustness, security, and capacity are directly linked to it. In the context of copy-protection, very low capacities are allowed as the goal of the detection is to determine the presence of the watermark, as opposed to a message carried by the watermark signal. A limit of one detection per 10 seconds of normally sampled audio (44100Hz) is common. On the other hand, security and robustness are crucial for the scheme. The asymmetric aspect of the scheme provides in itself good security: the detection is achieved without the watermark signal, and any brute attack attempting to subtract the watermark is almost impossible. Thus, our main concern is now the robustness to different types of attack, and in this context we use the following feature vector extraction method.

A fourier transform of length 1024 (size of one *frame*) 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 of the HAS. One component of $\mathbf{r}$ is extracted from each critical band cb and calculated as,

$$r_{cb} = \frac{1}{M} \sum_{i \in cb} \log_{10} |f_i|, \tag{3}$$

where M is the number of frequencies in the critical band. Using this extraction method, N/25 frames are required to extract a vector of length N. The vector $\mathbf{r}$ is finally obtained by filtering the extracted vector through a periodic function similar to that of quantisation process and which bounds its energy.

An interesting consequence of this extraction method is the psychoacoustic features of the scheme. Using the critical band method, we weight the watermark power according to the HAS, providing imperceptibility of the watermark presence. Secondly, since the watermark is duplicated on all frequencies in the critical band, there is some built-in robustness to desyncronisation, which is the main concern in this paper. We will see in section 5 that this feature vector choice provides a natural robustness to light desynchronisation, or imperfect synchronisation.

3. SYNCHRONISATION IN WATERMARKING

Desynchronisation attacks were first used in image watermarking with global affine transformations. In [7], Kutter *et al* proposed a method using sliding-correlation to estimate the translation and rotation parameters of a watermark pattern embedded in different known locations. The scheme was later defeated using random warping along with affine desynchronisation. In [8], Hartung *et al* approximated a global random distortion by a local affine transformation. In [9], Voloshynoskiy *et al* combined both methods and proposed a referenced watermark embedding at a local level. Although good results can be obtained with such techniques,

there are computationally very intensive and drastically increase the probability of false alarm.

In audio watermarking, a desynchronisation attack is usually some mapping function applied in the time-domain that can be classified as time-shift, linear time-scaling, random cropping, and non-linear time-scale modification (also referred to as warping [10]). In [11, 12], Malvar *et al* and Tachibana *et al* proposed methods to embed a watermark in the time-frequency plane of audio signals using a two-dimensional secret pseudo-random array. Repetition of the watermark pattern is used along with a sliding-correlator to achieve robustness against time and frequency desynchronisation. Besides the disadvantages associated with symmetric detection, these methods require an exhaustive search leading to large probability of false alarm. Moreover, they were defeated using a simple removal attack by exploiting the embedding redundancy and estimating the watermark locations and pattern. In [10], Eggers *et al* used a quantisation index modulation scheme to embed a pilot sequence in the time domain, and estimate the desynchronisation using a Viterbi tree search. Unfortunately, the complexity of such algorithms is known to increase rapidly with the dimension of the search space, and good performance of the technique is restricted to relatively smooth warping of the signal, well below the perception threshold of the HAS.

The HAS is very sensitive to distortion and imperceptible audio watermarking is generally achieved by shaping the watermark signal in the frequency domain using a fine psychoacoustic analysis of the original signal. In SSW the detection is commonly achieved by using vectors of frequency magnitudes. These vectors are usually obtained from FFT windows over the time-domain (a FFT window is the Fast Fourier Transform calculated over one frame of the audio sample). To ensure accurate detection, it is essential that the same FFT windows are chosen by the detection process as were chosen by the embedding process. The means of choosing the FFT windows along the signal is then an issue when a desynchronisation attack is carried out. Recently, Wu *et al* [13] and Kirovski *et al* [14] proposed a new type of synchronisation technique that relies on the extraction of robust features from the audio signal, such as beats or energy spikes, in order to locate the embedded information. This reduces, if not eliminates, the use of sliding-correlation in the detector. In this work, we propose to further develop these methods and combine them with very efficient detection schemes so as to provide increased security and robustness to a wide range of time-scale modifications.

Contrary to most watermarking schemes that take into account the problem of desyncronisation attacks, we have chosen to implement an asymmetric scheme in the sense that the watermark signal is not required at the detector. This important feature makes this scheme suitable for copy-protection applications.

4. FEATURE VECTOR EXTRACTION USING ROBUST PSYCHOACOUSTIC PROJECTION

A n^{th}-order side-informed watermarking scheme can be easily defeated by ensuring that the components of $\mathbf{r}$ are paired off incorrectly using a desynchronisation attack. This can be circumvented by using an extraction function $e(.)$ which uses robust psychoacoustic features of the host signal, referred to as robust salient points, to locate the extracted vector $\mathbf{r}$.

4.1. Use of Robust Salient Points

A simple way to view a desynchronisation attack is that the detector "doesn't look" at the right location in the studied signal. A solution to this problem is to devise a way of knowing the position used for the watermark embedding, and then to be able to trace these positions if they have been moved by a desynchronisation attack. In this context, salient points can be used as mobile reference points in the signal.

Wu $et\ al$ in [13] applied the concept of salient points to audio signal. In audio, peaks of energy are generally considered good salient points. Kalker $et\ al$ in [15] used a similar idea, the so-called "hashing for content identification", in which they extract certain features of the signal that make it uniquely identifiable. Although the basic ideas are similar, we require that a desynchronisation attack would be reflected on the position of our salient points (the ability to cope with temporal transformations). As such, the definition of our salient points will have to be done in the time domain, detailed in section 4.3, whereas Kalker $et\ al$ had the freedom to choose salient points by looking at the different characteristics of different domains, such as the frequency domain.

The psychoacoustic projection can be considered as a projection from the signal space to the embedding space. The dimension of the obtained embedding space is much smaller then the initial signal space, and therefore decreases dramatically the potential searches for the watermark.

4.2. Robust Salient Point Properties

It is the robustness to distortion of salient points that makes them ideally suited for this application as the points chosen at the embedding stage must be chosen again at the detection stage, even in the case of an attack. Hence, the psychoacoustic extraction must resist all the potential watermark attacks such as AWGN (Additive White Gaussian Noise), compression, filtering, resampling, etc. Another more specific attack might attempt to remove the salient points themselves. Faced with these different types of attacks, the salient points are chosen in sensitive regions so that they are not easily removed. Sensitive regions are generally areas of high energy of the signal, and carry mean-

ingful information according to the HAS. Consequently, removing such points would significantly degrade the quality of the signal. The definition of salient point regions is detailed in section 4.3.

Recently, in [14], Kirovski used periodic beats as salient points. However this approach is not applicable to many audio signals such as speech or jazz. In developing our extraction function, we have been mindful in ensuring that it is not signal dependent.

4.3. Psychoacoustic Projection Insight

In this section we describe the practical functioning of the psychoacoustic projection. Section 5.1 illustrates, with experimental results, the quality of the salient point extraction method.

The audio signal s is split into N macro-windows Typically the size of a macro-window varies between 5 and 15 frames where a frame is of length 1024 samples from an audio file sampled at 44100Hz. The salient point extraction function is applied over each macro-window m_i ($i \in$[1,N]) and a set of salient points $\{\phi_j\}$ is returned. From $\{\phi_j\}$, one salient point ϕ_p is selectively chosen. The embedding of the watermark is done with respect to ϕ_p, by using the frame [ϕ_p+1, ϕ_p+1024] to contribute components to the vector **r**. Recall that one frame of the signal gives 25 **r** components; the vector **r** is therefore built from many salient points. Consequently, if a salient point is incorrectly identified at detection, it will only minimally affect the detector decision as a whole.

A more detailed description of the salient point extraction function follows. The extraction function considers each sample, n, of the macro-window and computes three associated values for each of these: $\xi_1(n)$, $\xi_2(n)$, and $\rho(n)$. $\xi_1(n)$ and $\xi_2(n)$ are defined as a measure of the energy of η samples before and after the point n:

$$\xi_1(n) = \sum_{k=n-\eta+1}^{n} s(k)^2, \quad \xi_2(n) = \sum_{k=n+1}^{n+\eta} s(k)^2$$

The complexity of this computation is reduced to O(n) by calculating $\xi(n)$ as a function of $\xi(n-1)$.

The ratio ρ is then defined as:

$$\rho(n) = \frac{\xi_2(n)}{\xi_1(n)}.$$

The salient point extraction function selects points from the audio signal that are of both a high energy area and a high energy variation area. To do so, $\xi_2(n)$ and $\rho(n)$ are compared respectively to two thresholds Th_1, the energy threshold, and Th_2, the ratio threshold. If both tests are successful, the sample position is kept as a potential candidate salient point. In practice, we observed that potential candidates were organised in groups of consecutive sample points. So as to achieve greater stability for salient points, we then introduced a third threshold Th_3 which defines the minimum number of consecutive potential candidate points proceeding the candidate point under consideration. If this test is also successful, the position of the candidate sample point, n, is recorded as a salient point. Finally, a list of salient points is obtained for each macro-window, and from this list, one salient point is selectively chosen as the position for the watermark embedding. The criteria used in choosing the salient point ϕ_p from the set $\{\phi_j\}$ allows us to introduce an additional level of security.

There are a number of issues relating to the choice of the thresholds Th_1, Th_2 and Th_3. If these thresholds are set too high for the sample of audio being examined, no salient points may be found, in which case the embedding position will be a predefined default position known to the embedder and the detector. However, this default position has a much higher probability of failing after a desynchronisation attack. Conversely, if the thresholds are set too low, the number of salient points will be high with a greater risk of vulnerability to attacks, and a greater probability of error in the selection of ϕ_p from $\{\phi_j\}$. In future works, we will propose a systematic approach based on a statistical pre-processing of the audio sample in order to set optimised values for the thresholds. In section 5 the results are obtained with values for Th_1, Th_2 and Th_3 which provide an average of 10 salient points per macro-window.

4.4. Conclusion

The psychoacoustic projection dramatically reduces the size of the extraction/embedding space, and therefore minimizes the problem of the large exhaustive searches encountered in [11, 12].

The detection of the watermark relies on the ability to trace the displacement of the salient points caused by the desynchronisation attack.

5. EXPERIMENTAL RESULTS

This section presents experimental results carried out on real audio samples. A set of over 400 `wav` files, each 30 seconds long and sampled at 44100Hz, containing music from different artists and genres. The signal to watermark ratio (SWR) is set to 23dB for all the simulations, which ensures inaudibility of the watermark presence.

5.1. Impact of Imperfect Synchronisation on n^{th}-order detectors

In order to define the required effectiveness of the psychoacoustic projection at retrieving salient points, we first need to evaluate the impact of imperfect synchronisation on our n^{th} order decoders. In practice, it is not possible for the extraction of the salient points to be exact and therefore a predefined level of approximation must be tolerated.

In the following simulation we recorded the set of salient point positions used at the embedding $U_{embedding}$, and we then added a "jitter" jit to them in order to obtain the position of the salient points used at the detection $U_{detection}$:

$$U_{detection} = U_{embedding} + jit$$

where jit is a vector of integers uniformly distributed, with mean equal to zero, and spread over the interval [-ν ; ν], ν being the tolerance factor.

Fig.1 shows that, for a value of ν up to 50, the pdfs (probability density functions) of the correlation results under hypothesis H_0 and H_1, see section 2.1, are still sufficiently separated. This is explained by the fact that the watermarking scheme uses a frequency embedding that is spread widely and therefore benefits from the invariance by translation of the magnitude of the Fourier coefficient. This trait is essential for the use of the psychoacoustic extraction as it allows a tolerance to imperfect synchronisation.

The value of ν=50 corresponds to a significant loss of efficiency, even though still viable. We will set the tolerance factor ν at a lower level for better watermark detection by evaluating the performances of the projection according to the following rule: the position of the retrieved salient points must not differ from the original salient point positions by more than 20 samples.

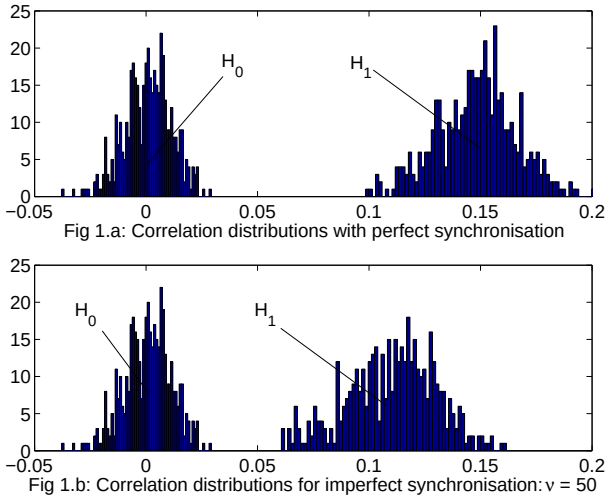

Fig. 1. Effect of imperfect synchronisation.

5.2. Robust Psychoacoustic Projection Performances

In this section we examine the performances of the psychoacoustic model in environments with a high level of noise. The attack used to test the accuracy of the projection is the classic AWGN attack. The range of the signal to noise ratio (SNR) used to illustrate the performances of the projection is from 50 dB to 10 dB. Audibility tests of such an attack have been conducted and all lab members agreed to set 40 dB as the threshold beyond which the noise addition is clearly audible. Fig.2 shows that the psychoacoustic projection performs well, even beyond the audibility threshold of 40 dB. Two tolerance factors, see section 5.1, are depicted, demonstrating a greater precision on the accuracy of the extraction function. Salient points are considered correct if their retrieved position differs from their original position by less then ν samples.

This experiment has been carried out on a set of 400 audio files which corresponds to the extraction of over 120000 salient points. For a noise addition corresponding to 40 dB, the limit of the audibility of the attack, 99.4$\%$ of the salient points are correctly retrieved for both tolerances. For a noise level of 23 dB below the signal, which considerable degrades the signal, 91.7$\%$ of the salient point positions retrieved are distant from their initial position by less the 4 samples, and 94$\%$ by less than 20 samples.

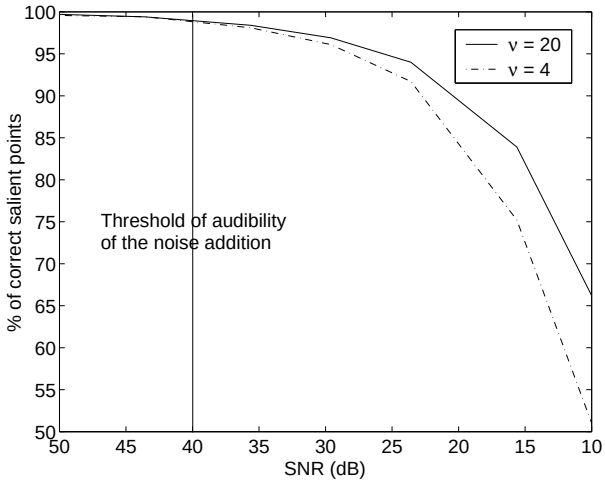

Fig. 2. Robustness of salient points.

5.3. Improvements to Desynchronisation Attacks

This last section shows the improvement of the complete watermarking scheme thanks to the psychoacoustic projection. Two types of desynchronisation attacks are considered here: random time-shift and random cropping. The random cropping is acknowledged by the watermarking community

as a major synchronisation issue. The first two sets of simulations in this section has been carried out with a 4^{th} order decoder. Finally, the ROC (Receiver Operating Characteristics) curve is given for 2^{nd}, and 4^{th} orders.

Fig.3 shows the resistance of the watermarking scheme to time-shift modifications with and without the psychoacoustic projection. The pdfs of the correlation results under hypothesis H_0 and H_1 are depicted after a random-shift attack. Shifts between 200 and 500 samples has been used in these simulations. Fig 3.a shows the vulnerability of the scheme to this type of transformation, and Fig 3.b illustrates the improvement brought on by the use of the psychoacoustic projection.

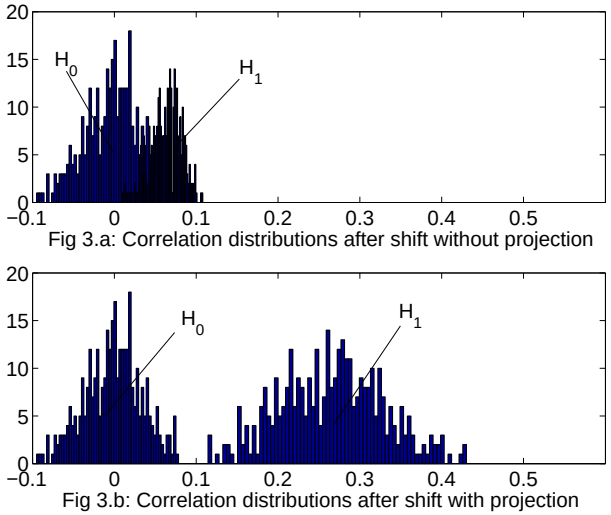

Fig. 3. Robustness to random shift.

In Fig.4, we look into the improvements with respect to the random cropping attack. We recall here that a cropping attack is achieved by randomly deleting samples from the original signal. In Fig.4 the impact of the use of the psychoacoustic projection is depicted in the context of random cropping. For this simulation the cropping rate has been taken between 500 and 1000 samples. Fig.4.a shows the impact of the attack on the computed correlations. Comparing Fig.3.a to Fig.4.a, it may seem that the random shift attack is more efficient than the random cropping attack. However it is important to realise that a random shift attack is relatively easy to inverse as the only parameter at stake is the offset used by the attacker. The random cropping case is much more complicated as the attacker has many degrees of freedom to implement the random removal of samples. The fact that Fig.4.a does not show catastrophic consequences after cropping is because we chose a uniformly distributed cropping. As such, the beginning of the signal is only slightly distorted. Moreover, we saw in section 5.1 that an imperfect synchronisation factor ν of 50 samples can be tolerated; it renders the desynchronisation effect acceptable on the first frames of the signal. Fig.4.b illustrates that the distortion introduced by deleting samples has been traced by the projection, ensuring a good retrieval of the FFT windows used for the watermark embedding. Once again, we observe well separated pdfs under hypothesis H_0 and H_1.

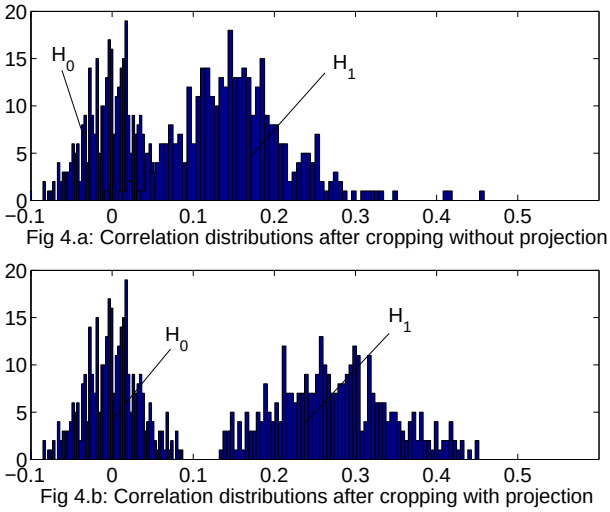

Fig. 4. Robustness to random cropping.

Fig.5 depicts the ROC curves, in the context of the random cropping attack, with 2^{nd} and 4^{th} order decoders.

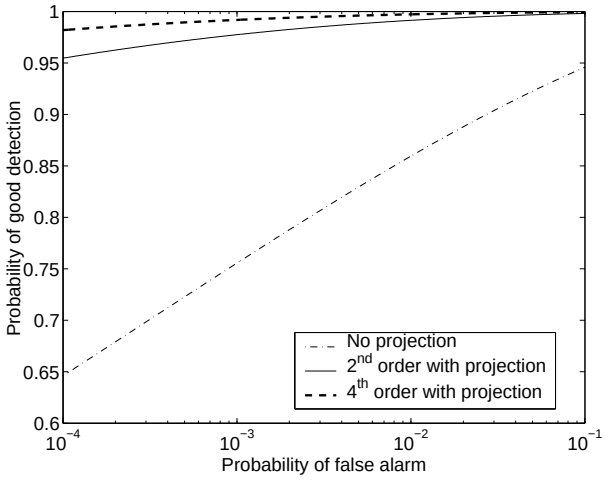

Fig. 5. ROC curves with random cropping.

6. CONCLUSION AND FUTURE WORK

In this paper, a new robust psychoacoustic projection is used to locate the extracted feature vectors as a means of synchronisation for side-informed n^{th}-order watermarking. The

new scheme benefits from the higher level of security gained from asymmetric detection, as well as the large efficiency of n^{th}-order detectors. Simulations were carried out on real audio signals under various desynchronisation and additive noise conditions, and they show that the psychoacoustic projection significantly improves the robustness of the watermarking scheme.

In future work, we will continue to develop the extraction function by including pre-processing of the signal so as to compute optimised values for the thresholds Th_1, Th_2, Th_3.

ACKNOWLEDGEMENTS

This project is supported by Enterprise Ireland Strategic Research Grant ST/2000/107/Y.

7. REFERENCES

[1] S. Voloshynovsky, S. Pereira, T. Pun, J.K. Su, and J.J. Eggers, "Attacks on digital watermarks: Classification, estimation based attacks and benchmarks," in *IEEE Communications Magazine*, Aug. 2001.

[2] N.J Hurley and G.C.M. Silvestre, "n^{th} order audio watermarking," in *Security and Watermarking of Multimedia Contents IV, Proc. of SPIE'02*, Jan. 2002.

[3] T. Furon, G.C.M. Silvestre, and N.J. Hurley, "Janis: Just another n^{th} order side-informed scheme," in *IEEE Proc. of ICIP'02*, Oct. 2002.

[4] G.C.M Silvestre, N.J. Hurley, and T. Furon, "Robustness and efficiency of non-linear side-informed watermarking," in *Proc. of Information Hiding'02*, Oct. 2002.

[5] T. Furon, I. Venturini, and P. Duhamel, "Unified approach of asymmetric watermarking schemes," in *Security and Watermarking of Multimedia Contents III, Proc. of SPIE'01*, Jan. 2001.

[6] M. Miller, I. Cox, and J. Bloom, "Informed embedding: exploiting image and detector information during watermark insertion," in *IEEE Proc. of ICIP'00*, Sept. 2000.

[7] M. Kutter, "Watermarking resisting to translation, rotation, and scaling," in *Proc. of SPIE'98*, Jan. 1998.

[8] F. Hartung, J. Su, and B. Girod, "Spread spectrum watermarking: Malicious attacks and counterattacks," in *Security and Watermarking of Multimedia Contents, Proc. SPIE'99*, Jan. 1999.

[9] S. Voloshynovskiy, F. Deguillaume, and T. Pun, "Multi-bit digital watermarking robust against local nonlinear geometrical distortions," in *IEEE Proc. of ICIP'01*, Oct. 2001.

[10] J.J. Eggers, R. Baulm, R. Tzschoppe, and J. Huber, "A channel model for watermarks subject to desynchronization attacks," in *Security and Watermarking of Multimedia Contents IV, Proc. of SPIE'02*, Jan. 2002.

[11] E. Malvar and D. Kirovski, "Robust covert communication over a public audio channel using spread spectrum," in *Proc. Information Hiding'01*, 2001.

[12] R. Tachibana, S. Shimizu, T. Nakamura, and S. Kobayashi, "An audio watermarking method robust against time and frequency fluctuation," in *Security and Watermarking of Multimedia Contents III, Proc. of SPIE'01*, Jan. 2001.

[13] C. Wu, P. Su, and C. Kuo, "A robust and efficient digital audio watermarking using audio content analysis," in *Security and Watermarking of Multimedia Contents III, Proc. of SPIE'01*, Jan. 2001.

[14] D. Kirovski and H. Attias, "Audio watermark robustness to desynchronization via beat detection," in *Proc. of Information Hiding'02*, Oct. 2002.

[15] J. Oostveen J. Haitsma, T. Kalker, "Robust audio hashing for content identification," in *Proceedings of the Content-Based Multimedia Indexing,2001*, 2001.