Chapter IV
Masking Models and Watermarking: A Discussion on Methods and Effectiveness

Mirko Luca Lobina
University of Cagliari, Italy

Luigi Atzori
University of Cagliari, Italy

Davide Mula
University Chair of Juridicual Informatics at LUISS Guido Carli, Italy

ABSTRACT

Many audio watermarking techniques presented in the last years make use of masking and psychological models derived from signal processing. Such a basic idea is winning because it guarantees a high level of robustness and bandwidth of the watermark as well as fidelity of the watermarked signal. This chapter first describes the relationship between digital right management, intellectual property, and use of watermarking techniques. Then, the crossing use of watermarking and masking models is detailed, providing schemes, examples, and references. Finally, the authors present two strategies that make use of a masking model, applied to a classic watermarking technique. The joint use of classic frameworks and masking models seems to be one of the trends for the future of research in watermarking. Several tests on the proposed strategies with the state of the art are also offered to give an idea of how to assess the effectiveness of a watermarking technique.

INTRODUCTION

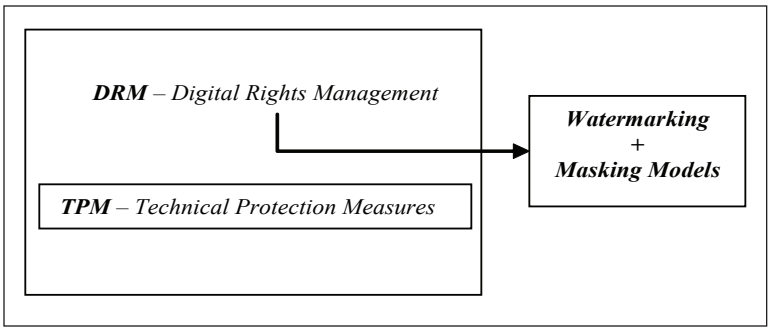
Ownership is one's right to possess something and decide what is to be done with it. The concept of ownership is distinct from possession, intended as having something in one's custody. Such demarcation, very easy to understand and prove in case of a material good (e.g., as a house or a car),

Copyright © 2008, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.

becomes particularly ephemeral in case of digital data. First of all, the diffusion and movement of digital content cannot be practically traced: as an example a collection of songs can be published on a Web site, a CD, or a DVD, and once it is legally distributed for the first time, there is no way to control its following circulation. Then, there are many ways to reproduce digital content and obtain copies identical to the original. Starting from a copy, a malicious person can alter or manipulate the original content and present himself/herself as the owner. This can result in great damage for the rights' genuine owner since an image or a song in a digital format are esteemed on the whole respect to their origin, which is a proof of quality, source, and content integrity. From this introduction it seems, and it is definitely true, that the encountering problems in ownership for digital content is the digital nature itself of the content: an image, printed on photopaper, can be copied with more difficulty without the original matrix (e.g., the negative, in case of a photo; or the master registration in case of an audio sample) and, in every case, the copy can be easily distinguished from the original. Furthermore, a photopaper image could not be present on millions of instances at the same time as happens to an image published on the Web. For all these reasons, different approaches have been developed for proving an origin's copyright information of digital content. On the whole, digital rights management (DRM) is any of these technologies when used to control access to digital data. DRM is not a technical protection measure (TPM). Actually, such technology controls or restricts the use and access of digital media content on electronic devices with such technologies installed, acting as components of a *DRM* design (See Figure 1). The objective of this chapter is to provide a comprehensive explanation of the crossing use of masking threshold and *watermarking* as one of the most effective *DRM* strategies.

This chapter is organized as follows. The first section provides an overview of *DRM*, discusses the distinction between cryptography and *watermarking* (explaining the relevant applicative fields in a *DRM* context of both solutions), describes *watermarking* in general, and introduces the *masking model*s. The second part is a description of the crossing use of the *masking model*s and *watermarking*. The third presents two works from the authors of audio *watermarking* schemes based on the use of the psychological masking threshold. Finally, some conclusions on the trade-off robustness/fidelity, and more generally on the use of *masking model*s in watermark schemes, are drawn.

Figure 1. DRM and TPM



BACKGROUND

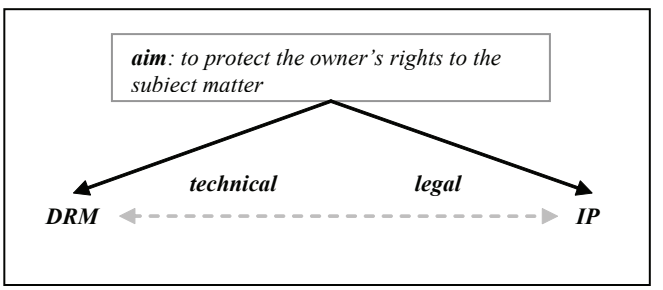
Digital Rights Management (DRM) and Intellectual Property (IP)

As already stated, digital rights management (DRM) is any technology used to control access to digital data. Such a term has been created to individuate the whole set of technical measures, where applied to digital media, or analog media released in digital format. The term "digital" is here repeated many times to underline that, while quality diminishes every copy with analog media, digital media may be copied a limitless number of times without degradation in the quality. That is, there is no difference in quality between the "master" and "copies." The last years have been characterized by a large pressure for legislation and regulation of DRM schemes' usage (specifically to publish the detailed functionalities of some of the most diffused DRM frameworks). An example of this will of regulation is the Digital Millennium Copyright Act (DMCA). That is, a United States copyright law that criminalizes creation and broadcasting of technology that can avoid measures taken to protect the owner's rights. An early example of *DRM* is the content scrambling system (i.e., CSS), based on a simple encryption algorithm (this last term is clearly explained when cryptography is introduced in the following). CSS has been used in DVDs since 1996 and requires DVD device manufacturers to sign a license agreement restricting the inclusion of certain features in their players that could be used to extract a high-quality digital copy disk. Approaching the problem of protecting the owner's rights from a legal point of view, we meet the term intellectual property (IP). IP defines how the holder of a legal entitlement (respect to ideas/information) can exercise his rights to the "subject matter" of the IP. Where, the term "subject matter" is used to reflect the idea that it is the product of the mind or the intellect. Thus, DRM and IP are complementary aspects of the same context (See Figure 2). DRM approaches the problem of protecting copyrights from a technical point of view. The legal basis of DRM is linked to IP and the underlying idea that the subject matter is the product of the mind or the intellect.

Cryptography and Watermarking

Cryptography and watermarking can be effective solutions for the distribution problem and rights management. In the following, the featuring factors of both techniques will be analyzed and discussed. Cryptography is the study of converting information from its comprehensible form into an incomprehensible format, rendering it unreadable without a secret knowledge (i.e., usually a numeric key). The field of cryptography has mainly expanded in two ways. The first one comprises

Figure 2. DRM and IP



mechanisms for more than just keeping secrets as schemes for digital signatures and digital cash. Cryptography finds a second application in secure protocols, used in infrastructure for computing and telecommunications. In many cases, such protocols are applied without the users being aware of their presence. Cryptography is concerned more with this second field, and it is used mainly in tenable communications. De facto, once digital data is widely distributed also via a protected channel (Macq & Quisquater, 1995), cryptography does not preclude its further storage or illegal actions such as copying. Digital *watermarking*, also referred to as simply *watermarking*, is an adaptation of paper watermarks to the digital milieu. That is, a pattern of bits is inserted into a digital content, identifying the origin's copyright information. Like printed watermarks, the purpose is to provide copyright protection for intellectual property in digital format. This implies robustness and capacity of the *watermarking* technique. Robustness measures the digital watermark ability to resist intentional/unintentional attacks, while capacity reflects the ability of conveying useful information embedded in the original signal. But unlike printed watermarks, which are intended to be somewhat visible, digital watermarks are designed to be completely, or at most partially, inaudible/invisible. This feature is referred to as fidelity, that is, a measure of the similarity (or of the antonym distance) between original and watermarked signal. Transparency implicitly means that fidelity has been attained.

From what has been said, it is clear that cryptography and watermarking are complementary aspects and solve different classes of problems with unlike approaches. On one hand, there is cryptography that provides more than simple content encryption/decryption (i.e., message/content confidential), but it is also involved in the distribution of a secret key in a secure manner to all the involved parties (i.e., secure channel). In this sense, cryptography provides the necessary protocols for communicating parties to steadily authenticate themselves with each other, and can be considered as an "active" solution to global security. In addition, the use of digital signatures and message authentication codes provides a means for verifying that the content has not been manipulated in any way (See Figure 3). While on the other hand, watermarking for copyright protection is useful for tracking illegally copied content, although this is a very "reactive" approach to content security.

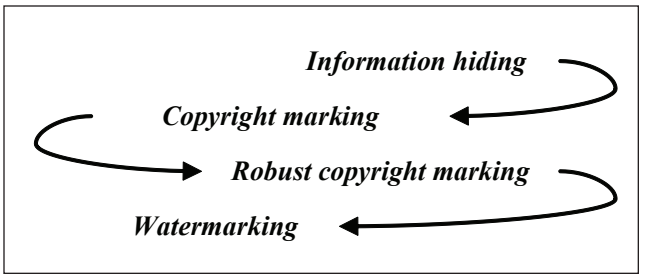
Watermarking in Detail

The first aim of this part is to provide the reader a very general overview of what is watermarking. Petiticolas, Anderson, and Kuhn, in their survey (1999), published at the end of the nineties on information hiding, introduce watermarking as a subcategory of copyright marking (See Figure 4). This is a very important account since, with this classification, they consider the robustness against possible attacks as one of the additional and mandatory requirements.

Figure 3. Cryptography and watermarking

Strategy:	***Applied to:***	***Typology:***
Cryptography	*Content/Channel*	*Active*
Watermarking	*Content*	*Reactive*

Figure 4. Information hiding and watermarking



Really, watermarking has three main means: monitoring (i.e., tracing of illegal copies), copyright protection (i.e., authentication of the copyright owner by the knowledge of the secret key to read the secret watermark), fingerprinting (i.e., during the phase of authentication to a secure network). Information carrier and indication of content manipulation will be discussed in detail in the following. All these possible applications are included in two main possible watermarking schemes: secret and public. Secret watermarking is used as authentication and content integrity mechanisms. This implies that the watermark is readable/understandable only by authorized persons with the knowledge about the existence/position of the embedded secret. On the other hand, public *watermarking* means that the watermark is just an information carrier.

As it has been introduced before, *watermarking* is the insertion of a pattern of bits identifying the origin's copyright information into a signal. The righteousness of the *watermarking* technique is measured in terms of robustness, capacity, transparency, and fidelity (See Figure 5).

A good *watermarking* technique should achieve these features as much as possible. Let us see briefly, in the following, the implications of such attributes:

- The host signal should be nondegraded and the embedded data be minimally perceptible; that is, the embedded data has to remain hidden to an observer of the watermarked signal.
- The host signal and the data to be embedded must be coherent in terms of embedding dominio. That is, the *watermarking* must be inserted, taking into account the features of the original signal (as compression or encoding). In this sense, it is very important to retrieve as much information as possible on the signal to be watermarked and on its use.
- The embedded data should be resistant to intentional/unintentional modifications. That is, the original signal is modified to try a removal of the *watermarking*. The possible modifications are wide and include chan-

Figure 5Watermarking features

Robustness: **measure of the digital watermark's ability to resist to intentional/unintentional attacks**
Capacity: **ability of conveying a number of useful information embedded in the original signal**
Fidelity: **distance between original and watermarked signals**
Transparency: **inaudibilityy/invisibility of the digital watermark**

nel noise, filtering, resampling, cropping, encoding, lossy compressing, printing and scanning, digital-to-analog (D/A) conversion, analog-to-digital (A/D) conversion, and so forth. Most of these will be described in detail when introducing the typologies of possible attacks.

Let us analyze the problem of nondegrading the original signal. Obviously, trade-offs exist between almost three factors: total amount of data that can be embedded (i.e., "bandwidth"), transparency respect to the host signal, and robustness/resistence to modifications. A technique can be planned that, constraining the host signal degradation, tries to reach high embedding data, or high robustness. The two last factors are mutually divergent: as one increases, the other must decrease (See Figure 6).

The previous statement cited three factors as being true for every possible configuration of such parameters. For this reason, different kinds of techniques have been developed for different applications and requirements (See Figure 7). That is, the application fixes the requirements, the three factors' trade-offs and consequently, the technique itself.

In Bender (Bender, Gruhl, Morimoto, & Lu, 1996), another classification of *watermarking* techniques is proposed, depending on the applica-

Figure 6. Robustness vs. bandwidth

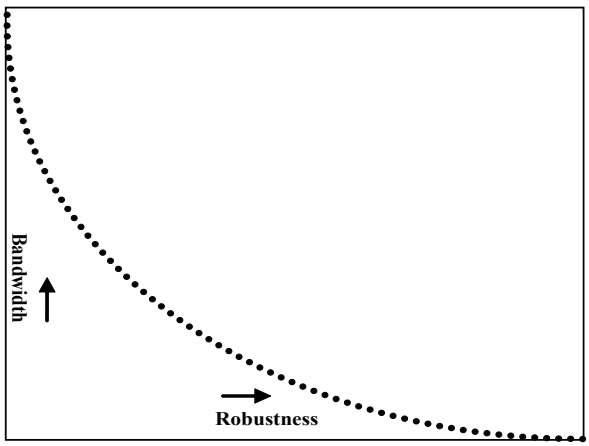
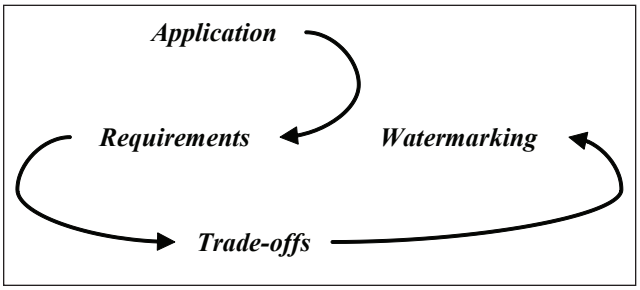


Figure 7. Requirements and watermarking's trade-offs



tion in terms of the amount of information that is embedded in the host signal. In this sense, the simpler technique implies the embedded data are used to place an indication of ownership in the host signal; the position is the copyright indication itself. A second category is the tamper-proof, while the third is the feature-location. Tamper-proof frameworks can be effectively included in two categories as "fragile" or "semifragile," as proposed in Kazakeviciute (Kazakeviciute, Januskevicius, Rosenbaum, & Schumann, 2005). Fragile means that watermarks act as a digital signature, allowing the detection of any modifications of the content. Instead, semifragile frameworks allow one to detect and localize any modification of the content changes. The second category is more effective respect to the set of functionalities from the digital signature. In Wu and Liu (1998), a set of desired features for tamper-proof frameworks is presented as follows:

- Determination whether a host signal has been altered or not.
- Integration of authentication data with a host signal rather than as a separate data file.
- Invisibility of the embedded authentication data under normal viewing/auditing conditions.
- Localization of the alteration that was made on the signal.
- Storage of the watermarked signal in lossy compression format.

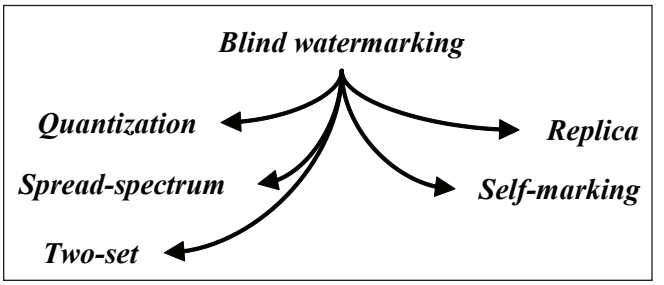
Feature-location watermarks enable one to identify individual content features by introducing a certain degree of modification in the host signal. Generally, it requires a large amount of information respect to the other two previous techniques. Always referring to Figure 7, the "application" level includes also the case the signal to be marked is an audio or an image signal. The aim here is to exploit the main traits in the methods, considering both contexts (i.e., audio/image signals) and low/high bandwidths (always refer to Figure 7).

Blind audio *watermarking* (Eggers, Su, & Girod, 2000) techniques are first described in detail as follows. The term blind is here used to indicate that the watermarking scheme can detect and extract watermarks without nonwatermarked/original audio signal. Obviously, the blind methods need an effective method for watermark's self-detection. Compared to nonblind techniques (which implies the original signal for comparing purposes), there is a saving of *50%* for storage and bandwidth.

To give the reader a comprehensive and simple overview, five categories (See Figure 8) of frameworks are described. Of course, much more techniques are accessible (take also into account that a watermarking method must be obligatory to be published). This section does not introduce methods based on psychological models since they will be described in detail in the following.

Wikipedia introduces the term quantization as the procedure of constraining something to

Figure 8. Blind-audio watermarking technique



a discrete set of values. In the specific field of audio signals, it is the process of approximating a continuous range of values (or a very large set of possible discrete values) by a relatively small set of discrete symbols or integer values. Thus, a quantization method quantizes a value of the original signal (let us say χ), and assigns it a new value based on the quantized sample value D and function $q(\chi, D)$, in a way $y = q(\chi, D) \pm D$, where the signs $\pm$ depend on the value of the value of the binary watermark message {0,1}. This scheme is quite robust (Chen & Wornell, 1999) and resistant to several types of attack. The detection method is exactly the inverse with respect to the embedding process.

Spread spectrum was originally developed for military applications. Since its declassification, it has found civilian use, particularly in communications for cellular telephony. Spread spectrum is one of the most diffused and discussed schemes in state of the art (Cox, Kilian, Leigton, Shamoon, 1996; Cvejic, Keskinarkaus, & Seppanen, 2001; Kim, 2000). Several implementations have been proposed in the past, but in this chapter, the basic idea is presented. In spread spectrum, a narrowband signal (the message to be transmitted) is modulated by a broadband carrier signal that spreads the original spectrum. Spread spectrum is a correlation method that inserts a pseudorandom sequence into the original signal, in time or transform domain, and detects the watermark by calculating the correlation between the watermarked signal and the pseudorandom sequence. The carrier is the pseudorandom sequence (i.e., with statistical properties similar to a truly random signal) obtained with a random-number generator that has been initialized with a particular seed, known only to the owner. Thus, a generic binary message (the watermark) can be modulated by the pseudorandom sequence generated by means of the seed. Obviously, it is not possible to determine the watermark without the knowledge of the seed. Generally, such a factor is the privileged information that the sender and an authorized receiver possesses. Apart form the seed, spread spectrum presents another element of robustness. That is, a large signal power is distributed over the entire frequency spectrum, so only a small amount of power is added at each frequency.

The two-set watermarking scheme marks two sets from the original signal differently. Thus, the presence of the difference itself between the two sets allows one to understand if a signal is marked or not. Also, in this case, the state of the art has been very rich and purposeful (Arnold, 2000; Bender et al., 1996; Yeo & Kim, 2003). Many variations have been proposed. In this section, we will be deep in analysis to a particular approach, that is, the patchwork approach, as in the following section "Future Trend," the crossing use of the audio *masking model* and two-set patchwork approach is described. Original patchwork approach works in embedding phase in two steps using a statistical approach: first of all, two sets, or patches, of the original signal are chosen in time/transform domain; then, a small constant value (let us say d) is added to one patch and subtracted from the other. Obviously, the patch in is composed by more than one element so that it is possible to evaluate the mean value of the elements composing each of the two patched sets (let say $\overline{\beta}^*$ and $\overline{\alpha}^*$). It is also possible to evaluate the mean value of the difference of the single mean values of the elements of the two patched sets (let say $E(\overline{\beta}^* - \overline{\alpha}^*)$[1]), while the corresponding mean value of the difference of the single mean values of the elements of the two original sets can be indicated with $E(\overline{\beta} - \overline{\alpha})$. This change in $E(\overline{\beta}^* - \overline{\alpha}^*)$ and $E(\overline{\beta} - \overline{\alpha})$ is very useful in the detection phase as the value itself changes from the original value $\overline{x}$ to that corresponding to the patched situation $\overline{x} + 2d$. Also in this case the position, (with respect to the original signal) of the two sets is chosen by means of a random number generator with a seed. The presence of the watermark can be individuated only with the exact knowledge of the seed using the statistical approach previously described. Note that the statistical distribution of the values of the two patches

does not change (because the same quantity is added and subtracted); it is the difference in the mean values that allows a detection (See Figure 9). The detection is performed by means of a hypothesis test. That is, given the null hypothesis (denoted H_o) against the alternative hypothesis (denoted H_1), the outcome of a hypothesis test is "reject H_o" or "do not reject H_o." In this case, the null hypothesis can be formulated as "the signal is not watermarked." Thus, the aim is to reject the null hypothesis. The performance of the patchwork scheme strongly depends on the distance between $E(\overline{\beta}^* - \overline{\alpha}^*)$ and $E(\overline{\beta} - \overline{\alpha})$, which is function of *d*. Obviously, the greater the factor *d* corresponds to a lesser probability of failure in the detection phase, but also to a minus transparency of the watermark in the original signal. Thus, also in this case, the trade-off robustness vs. transparency must be carefully evaluated.

The idea behind replica method is that the original signal itself can be used as a watermark. Echo hiding (Bender et al., 1996) is a good example of a replica scheme. An echo is characterized by

Figure 9. Comparison between watermarked and un-watermarked distributions of the mean differences

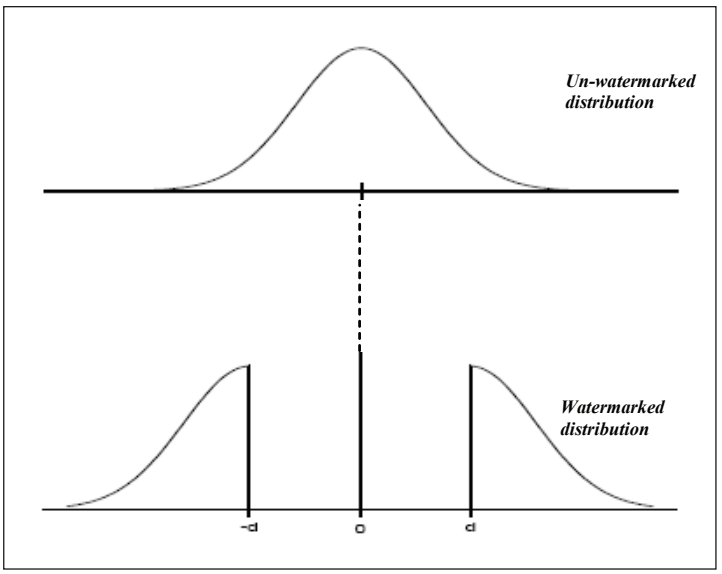
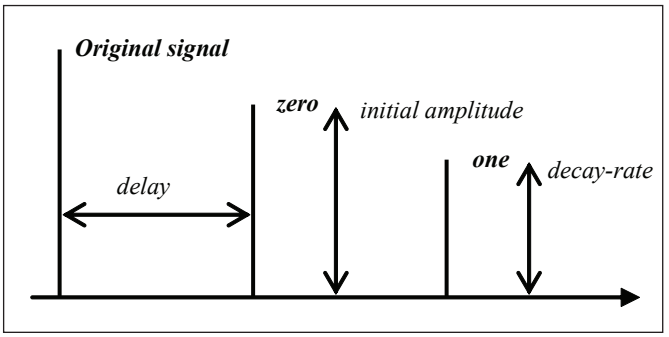


Figure 10. Replica method



three factors: initial amplitude α, delay (offset), and decay rate (See Figure 10).

The echo is perceived as added tone with respect to the original signal (See Figure 11).

The detection phase can be considered as a decoding process involving the evaluation of the cepstrum of the echoed signal. The "cepstrum" can be used to separate the excitation signal (which contains the pitch of an audio signal) and the transfer function (which contains the audio quality). Without offering too many technical aspects (see Bender et al., 1996) the basic idea behind detection in a replica scheme is to determine the spacing between the echoes.

Finally, the self-marking method (Mansour & Tewfik, 2001; Wu, Su, & Kuo, 2000) embeds a signal (a self-evident mark) in the time/frequency domain in a way that is easily noticeable. In this chapter, we report specifically for time-scale modification strategy (Mansour & Tewfik, 2001), which basically is the process of compressing/expanding the time-scale of the audio signal between salient points. In Mansour and Tewfik (2001), such points are selected extrema of the wavelet decomposition, and the length of the interval is chosen according to the watermark to be embedded. In the last decade, the study of audio watermarking has been characterized by a huge improvement. This first part surveys the main five categories of audio *watermarking* apart from the use of masking and psychological models. A series of general thoughts can be expressed on such methods. Quantization approach is simple, but at the same time not very robust to several kinds of attack. In this sense, it is better than echo hiding, which is more complex but presents the same lacks in robustness. On the other hand, we have robust schemes like the two-set (and specifically the patchwork strategy) and replica approaches. Finally, the spread spectrum, which is effective, but can lead to serious modifications of the original signal. For this reason, it requires a psychological model for inaudible watermark embedding. This last consideration can also be considered true for several of the proposed frameworks.

Figure 11. Outcome of the replica method

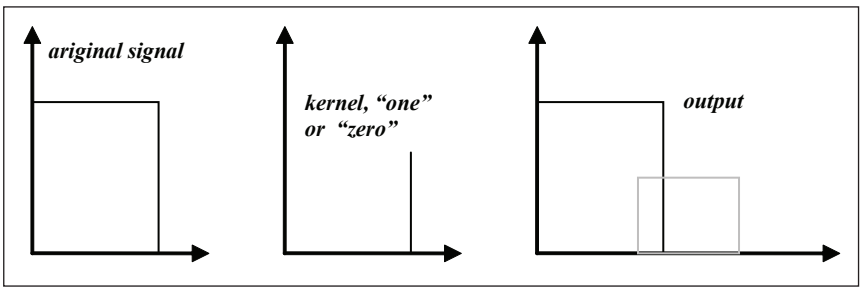
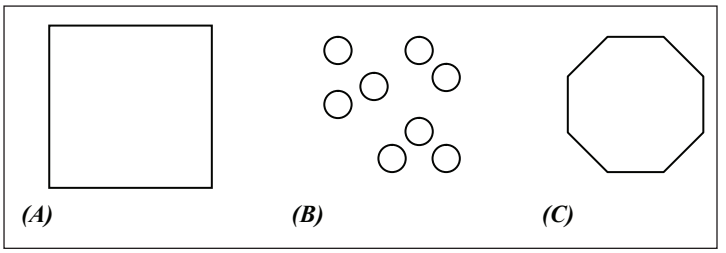


Figure 12. Patch shapes



Most of the described strategies can be applied both for images and audio signals. As an example, the patchwork strategy, a low bandwidth strategy, is applied on the same way for image and audio. The difference is about the importance of the patch shape (See Figure 12) and its visibility within the image. The optimal choice of patch shape depends upon the expected image modifications, that is, if a JPEG image is considered, it is preferable to use a patch that embeds the energy in low frequencies. A comprehensive description of such a relation is proposed in (Bender et al., 1996).

The patchwork strategy is obviously an example, but the main differences between implementations of watermarking frameworks of image and audio signals are about two factors. That is, on one hand the human visual/auditory systems (HVS/HAS), and on the other the nature of the original signal itself. In the following, a brief explanation of both such factors is proposed (Bender et al., 1996).

A still image (the original signal to be watermarked) is typically an 8-bit 200×200 pixel image. Furthermore, it is also realistic to suppose that an image will be subject to many kinds of operations, such as filtering, lossy compression, and blurring. At the same time, the HVS presents sensitivity to contrast as a function of spatial frequency and the masking effect of edges (both in luminance and chrominance). The HVS is not very sensible to small changes in luminance. This is the first "hole" of the HVS. The second is the HVS's insensitivity to very low spatial frequencies such as continuous changes in brightness across an image. Many texts can be cited that propose a comprehensive description of human eye and visual system. In this chapter, we quote from Tovèe (1996), Marr (1982), and Delaigle (Delaigle, Devleeschouwer, Macq, & Langendijk, 2002). High-quality audio is usually a 16-bit 44.1 KHz signal (as Windows Audio-Visual (WAV)). A compressed audio (i.e., from (ISO MPEG-AUDIO) perceptual encoding standard, originally described in Brandenburg and Stoll (1992), presents statistics of the signal drastically changed with respect to the original, while the characteristics that a listener perceives are the same. Many documents and much information can be retrieved about audio standards and HAS, but for a base knowledge the authors suggest one read the original standard (Brandenburg & Stoll, 1992; Pan, 1995), while the HAS will be described in the following sections.

Masking Models

Psychological and *masking model*s are usually inherited from data compression mechanisms and are very difficult to understand and explain in detail. Thus, the idea of the authors is to provide the reader with general key information and some references to find detailed information. First of all, independently from the nature of the signal itself, a masking model determines the levels of intensity (as a function of time or frequency) of a signal, to be perceived in presence of another signal. The masking model is usually a function or a set of data called simply, mask. The mask modulates the watermark to ensure its transparency, or to increase the watermark's energy for a given fidelity/transparency constraint. The main problem is specifically due to the fidelity feature; in fact, while concentrating on the increase of the trade-off energy/fidelity, little attention has been paid to the potential vulnerability of the masked watermark to a malicious attack. The scope of this chapter is to provide the reader a comprehensive description of watermark methods based on masking models, evaluate their effectiveness with respect to an intentional attack, and draw some conclusions for future trends. The concept of mask, as employed in *masking model*s, is used to signify a set of values (i.e., a matrix) that is computed on individual processing blocks (i.e., a block of 20 ms of an audio trace, or 8 × 8 pixel of an image) of the original signal. The original signal is defined as stationary in its temporal and spectral features, in an individual block. This is quite important for the definition of a correct mask. The reading of

the mask leads one to understand in which regions of the transform domain a signal can be suitably inserted in a way its energy remains lower than the energy-threshold fixed by the mask, and thus can be as inaudible as possible. That is, the mask shapes the message (i.e., the sequence of bits to be inserted in the original signal to prove its copyright) to obtain the watermark. The shaping could involve all, or a portion of the transform domain: the mask can be spread over the entire transform domain (i.e., an identity matrix, multiplied by a small constant), or can be selectively embedded in particular regions of the transform domain (i.e., a matrix with all null values except from specific regions with no-zero coefficients). The retrieval of the mask differs from audio to image samples. Let us consider first of all the audio context. The psychoacoustic model is based on many studies of human perception. These studies have shown how an average human responds to sound near in frequency and/or time. The main two properties of human perception are absolute threshold of hearing (ATH) and auditory masking.

ATH means that humans can hear frequencies in the range from 20 *Hz* to 20,000 *Hz*. However, this does not mean that all frequencies are heard in the same way. This consideration divides the range 20 *Hz* to 20,000 *Hz* in critical bands that are nonuniform, nonlinear, and dependent on the heard sound. Signals within one critical bandwidth are hard to separate for a human observer. Without introducing the relationship between Hertz and Bark (for a deeper investigation the authors quote Smith & Abel, 1999), ATH has been defined as a function of frequency, and represents the threshold of audibility of a sound at a certain frequency (see Figure 13).

The second property, auditory masking, means that humans do not have the ability to hear minute differences in frequency. This inability can be determined as the function of three factors: tone and noise maskers, and masking effect. A tone has a sharp peak in the frequency spectrum, which indicates that it is above the ambient noise of the signal. When a tone is very near to another in frequency, there is high probability that the first should mask the second (see Figure 14).

Humans have difficulty discerning signals within a critical band; the noise found within each of the bands can be combined to form one mask. Thus, the idea is to take all frequency components within a critical band that do not fit within tone

Figure 13. ATH, absolute threshold of hearing

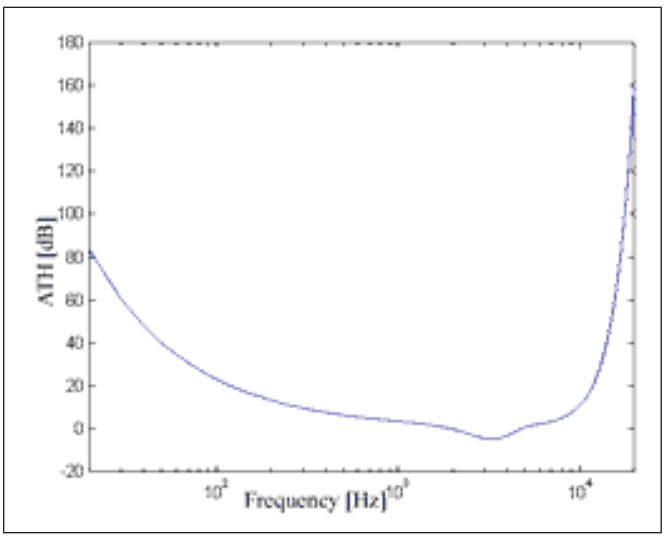
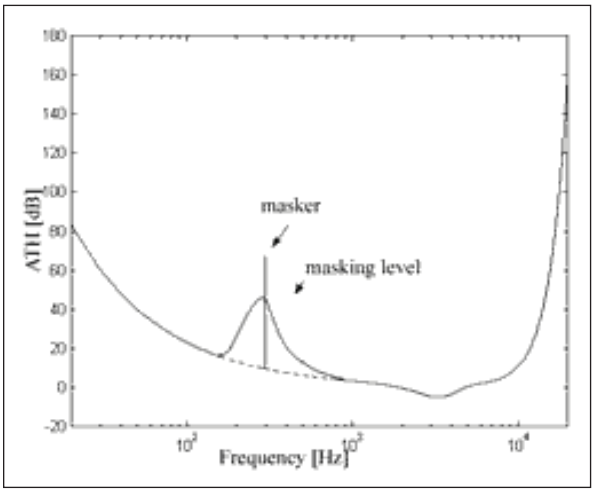


Figure 14. Frequency masking



neighbourhoods, add them together, and place them at the geometric mean location within the critical band. Finally, the masking effect is related to the fact that the maskers, previously determined, affect not only the frequencies within a critical band, but also in surrounding bands: this results in a spreading of the effects of the masker (tone/noise). Apart from this straightforward introduction, if there are multiple noise and tone maskers, the overall effect is a little harder to determine. Several references can be quoted to provide the reader an idea of how the audio *masking model*s and ATH work. First of all, the standard (ISO MPEG-AUDIO; Pan, 1995), which represents a good starting point also to nonspecialized people. Then, to understand the anatomy and functionalities of the HAS: (Probst, 1993; Robinson & Hawksford, 1999). Campbell and Kulikowski (1966) and Pantle and Sekuler (1969) presented masking effects in the visual domain in their works in late 1960s, and later in Tovèe (1996). Such effects, selective for bands of spatial frequencies and orientations, are the basis for the HVS and, as for the audio context, result in visual masking (Ferweda, Pattanaik, Shirley, & Greenberg, 1997). The reader can find an exhaustive description of the parameters for masking in Legge and Foley (1980). This last article is also the reference for many of the proposed image data hiding strategies presented in the last years. The authors also suggests Pennebaker and Mitchell (1993), Lambrecht and Verscheure (1996), and Lubin (1993) and an important study on appearance colour pattern (Poirson & Wandell, 1993).

Watermarking Framework Based on Masking Models

A consistent number of watermark frameworks have been developed that control the trade-off between robustness and inaudibility by means of masking models. The approach is similar for audio and image watermarking. As an example, the audio masking model is here shortly presented as previously stated; the spectral audio masking model is based on a very simple observation. Given a tone at a certain frequency, humans do not hear this tone, and all the others below/above it, if its intensity is below the energy-threshold fixed by the mask. The most widely used audio masking model can be detailed as follows: (1) divide the signal in 20 *ms* nonoverlapping blocks; and for each block: (2) compute short-term power spectrum; (3) evaluate noise and pure-tones compo-

nents; (4) remove all the masked components[2]; (5) compute the individual masking thresholds[3]; (6) compute the global masking thresholds. Usually the watermark is inserted 3-6 dB below the mask to enhance its inaudibility. In the following, we briefly describe the strategy presented in Mitchell et al. (Mitchell, Bin, Tewfik, & Boney, 1998) to give the reader an example of a well-projected *watermarking* framework based on *masking model*s. The macrosteps of the embedding phase (for a block of fixed dimension from the audio signal) are:

- Compute the power spectrum of the audio segment.
- Compute the frequency mask of the power spectrum.
- Use the mask to create the shaped author-signature.
- Compute the inverse FFT (IFFT) of the frequency shaped noise.
- Compute the temporal mask.
- Use the temporal mask to further shape the frequency shaped noise creating the watermark of that audio segment.
- Create the watermarked block.

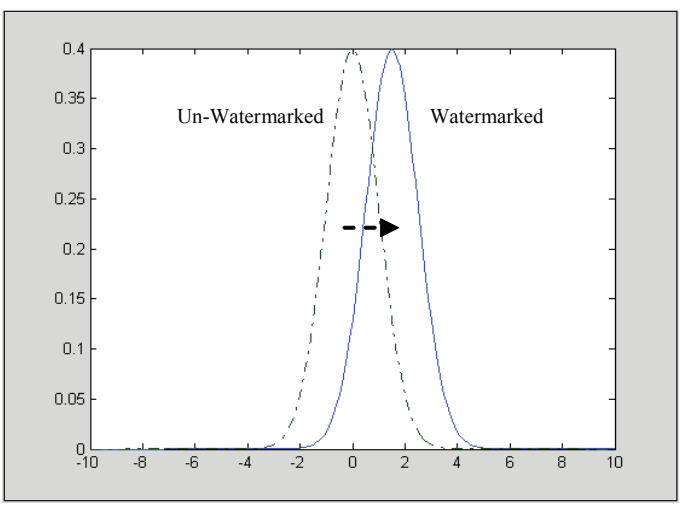
That is, the temporal/spectral masking models are applied (Steps 2/5) to create a signal-shaped watermark. We have described this strategy also because it is probably the most representative and simple to be valued for nonspecialized readers. The detection phase, without describing too many technical aspects that can be found in (Mitchell et al., 1998), is always based on a hypothesis test as described in a previous section. The basic idea at the base of detection phase is that, as the embedded watermark is noise-like, a pirate has insufficient knowledge to directly remove the watermark. The state of the art is very prosperous of techniques based on *masking model*s. In this chapter, we quote several papers both for the novelty of the presented strategy and the results and comparisons. The reader should examine in particular: (Herkiloglu, Sener, & Gunsel, 2004; Quan & Zhang, 2004; Robert & Picard, 2005; Swanson, Bin, & Tewfik, 1998; Tovèe, 1996; Wang, Xu, & Qian, 2005; Xie & Shen, 2005; Xing & Scordilis, 2005). The reader can find a better explanation of aspects related to the use of psychological (i.e., masking) models in watermarking in IBM Research CyberJournal (Craver, Memon, & Yeo, & Yeung, 1996; Craver, Memon, Yeo, & Yeung, 1997).

FUTURE TRENDS

The recent years have been characterized by a growing diffusion in the fruition of digital audio contents and consequently in the need for copyright and ownership property. The watermarking techniques represent a good solution for these supplies: a mark is opportunely inserted in a host signal in a way so that its ownership is provable. Lots of strategies have been presented in the recent past with this purpose. Several of these techniques inherited their core from image watermarking; more in general, this legacy was not always possible due to the differences in sensibility and perception between human ear and eye. A set of basic features for a reliable watermarking strategy was presented in Arnold (2000). Two characters are most significant and, apparently, contradicting: inaudibility and robustness to signal processing. Inaudibility means that the differences between the original and the watermarked signal should not be perceivable by the human ear. Second, the watermark must be robust against intentional or unintentional attacks. One of the most impairing attacks is the signal processing, specifically, lossy compression. Such a compression guarantees enhanced portability of digital information, but can have an undesirable effect on the embedded watermark. The development of our strategy is accomplished by referring constantly to these two features. In this section we propose two watermarking frameworks for audio contents,

based on the use of a psychoacoustic model. Both works are based on the patchwork strategy, originally described in Bender (1996), for image watermarking. Further adaptations have been proposed to improve its efficiency when applied to audio contents. Such changes are mainly based on the right dimensioning of the patch in order the introduced noise that should not be perceivable by the end-user. The core of the proposed technique is the shaping of the patch so that two optimality criteria (i.e., inaudibility and robustness to lossy compression) can be satisfied. The inaudibility is guaranteed by shaping the patch vs. the minimum masking threshold, obtained with the psychoacoustic model 2. For both the proposed frameworks, the original signal is not needed for the detection of the watermark. The first strategy is an adaptive approach of the patchwork algorithm. The patchwork is originally presented in Bender (1996) for image *watermarking*. The original implementation of this technique presents several limitations when applied to audio samples. Quite a few adaptations have been proposed to improve considerably its performance (Arnold, 2000; Bassia, Pitas, & Nicholaidis, 2001; Kim, 2003; Yeo et al., 2003). From these studies, the work-dominion and the adaptive patch shaping appear as the key points for the applying of the original strategy to audio samples such as for its perfection. The proposed strategy works on an assumption introduced in Bender (1996): treating patches of several points has the effect of shifting the noise to low frequencies, where there is lower probability to be filtered by lossy compression techniques. This concept has also been introduced previously in this chapter. The problem is the dimension of the patch. Such a factor is fixed by comparing the spectrum of watermarked signal to the minimum masking threshold, as obtained referring to the psychoacoustic model 2 (ISO MPEG-AUDIO). The patch shaping is performed in the Fourier dominion. The proposed technique is applied to audio samples and compared with the adaptive patchwork state of the art, referring to the framework proposed in Gordy and Bruton (2000). As previously described, the patchwork strategy is a two-set method, that is, it makes two different sets from a host signal (Yeo et al., 2003). This difference is used to verify, or not, a hypothesis H_o (e.g., the watermark is embedded).

Figure 15. Distribution of the mean difference of the samples in un-watermarked and watermarked signals



The original strategy (Bender, 1996) is applied to sets with more than 5,000 elements. The samples of each subset are considered uniformly distributed and with equal mean values. The elements are modified by adding and subtracting the same quantity *d*. Thus, the detection of a watermark is related to the condition $E[A_{marked} - B_{marked}] = 2d$. Several of these statements must be reassessed when working with audio samples (Arnold, 2000). In particular, the distribution of the sample value is assumed as normal (See Figure15). Recent approaches modify the original strategy to better take into account the human ear's sensibility to noise interferences. These methods can be classified in temporal and spectral approaches, depending on the domain where the watermark is embedded. In Bassia et al. (2001) a technique is proposed that is based on the transformation of time-domain data. A set of *N* samples, corresponding to 1*sec* of stereo audio signal, is modified by a watermark signal w(i). Arnold (2000), Kim (2003), and Yeo et al. (2003) propose spectral patchwork approaches. In particular, Arnold (2000) works with a dataset of 2*N* Fourier coefficients. The relationship between *d* and the elements of the dataset is multiplicative. The parameter *d* is adaptively chosen to prevent perceptual audibility, basing on the characteristics of the audio signal (i.e., it introduces for the first time the concept of power density function in the hypothesis tests). In Kim (2003), the patchwork algorithm is applied to the coarsest wavelet coefficients, providing a fast synchronization between the watermark embedding and detection. While in Yeo et al. (2003), the modified patchwork algorithm (MPA) is presented. Such an approach is very robust due to three attributes: the factor *d* is evaluated adaptively and is based on sample mean and variance; the patch size in the transformed domain is very little: this guarantees good inaudibility; finally, a sign function is used to enhance the detection rate. These features are included in an embedding function so that the distance between the sample means of the two sets is bigger than a certain value d. The temporal approaches are easier to implement than the spectral ones; at the same time, they present several weaknesses against general signal processing modifications (Kim, 2003). The association between a watermarking algorithm and a noisy communication system is not new (Muntean, Grivel, Nafornita, Najim, 2002). Actually, a watermarking strategy adds a mark (i.e., the noise) in a host signal (i.e., the communication channel). In this sense, the watermark embedding can be considered as an operation of channel coding: the watermark is adapted to the characteristics of the transmission channel (i.e., the host signal in which the watermark should be embedded). In case of audio contents, where it is usually considered as an impairment, the sensibility of the human ear can be used as a way to spread and dimension the watermark. The human auditory system (HAS) is well known, that is, it is sensitive to specific frequencies (i.e., from 2*KHz* to 4*KHz*) and reacts to specific events (i.e., frequency and temporal masking). Given a signal *S*, it is possible to recover its minimum masking threshold. The minimum masking threshold of audibility represents a limit between the audible and inaudible signals for *S* at different frequencies. Independently from *S*, it is also possible to recover the absolute threshold of hearing (ATH). Such a curve (i.e., referred to as quiet curve (Cvejic et al., 2001) is different than the previous and defines the required intensity of a single sound expressed in unit of decibel (dB) to be heard in the absence of another sound (Zwicker & Zwicker, 1991). Several methods outside the patchwork fashion have been proposed that make use of psychoacoustic models to guarantee perceptual inaudibility of the mark (Arnold, 2002; Boney, Twefik, & Hamdy, 1996; Cvejic et al., 2001). Usually, the state-of-the-art methods shape the watermark referring mainly to the quiet. The filtered watermark signal is scaled in order to embed the watermark noise below the quiet curve (ISO,1993). In addition, other methods increase the noise energy of the watermark, referring undeniably to the minimum threshold

of audibility. Such a threshold can be recovered through a well-defined psychoacoustic model. The MPEG/audio standard provides two example implementations of the psychoacoustic model. Psychoacoustic model 1 is less complex than psychoacoustic model 2 and has more compromises to simplify the calculations. Either model works for any of the layers of compression. However, only model 2 includes specific modifications to accommodate Layer III. In this chapter, we refer to model 2 differently from the past approaches.

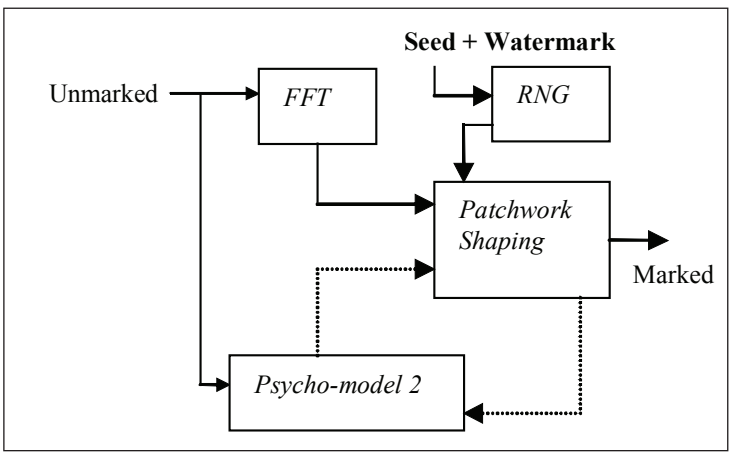
As already stated, the proposed patchwork strategy modifies two sets of N elements/coefficients from the original signal (i.e., signalUn-Marked). The signalMarked strongly belongs to the correspondent signalUn-Marked. The core of our strategy is the shaping of the frequency-response of the mark signal using psychoacoustic model 2. The algorithm proposed in this chapter embeds the watermark in the frequency domain by modifying $2N$ Fourier coefficients. The choice of this transform domain is justified by the use of the psychoacoustic model. The embedding steps (see Figure 16) can be summarized as follows:

1. Evaluate the threshold of minimum audibility for the signalUn-Marked, referring to psychoacoustic model 2.
2. Map the secret key and the watermark to the seed of a random number generator. Next, generate two N-points index sets $I_N^A = \{a_1, a_2, ..., a_N\}$ and $I_N^B = \{b_1, b_2, ..., b_N\}$.
3. Let $X = \{X_1, X_2, ..., X_{2N}\}$ be 2N DFT coefficients of the signalUn-Marked, corresponding to the index sets I_N^A and I_N^B.
4. The original amplitude of the patch and the number of retouched coefficients, starting from the generic elements of index a_i or b_i, have respectively standard values (δ, θ). Such values are modified iteratively to verify that the spectrum of the watermark signal is under the minimum audibility threshold (i.e., obtained from point 1). Iteratively means a constant referring to the block of model 2 from the block shaping (see the dotted loop in Figure 16).
5. The time-domain representation of the output signal is found, applying an Inverse DFT to the signalMarked .

The phase of detection is as follows:

1. Define two test hypotheses: H_o (the watermark in not embedded) and H_1 (the watermark is embedded).

Figure 16. Steps (1-4) of the Patchwork shaping algorithm



2. Map the seed and the watermark to a random number generator and generate two sets $I_N^{A'}$ and $I_N^{B'}$.
3. Fix a threshold Δ for the detection, and evaluate the mean value (i.e., $\bar{\cdot} = E(\cdot)$) of the random variable $z = a_i' - b_i'$, for $a_i' \in \{I_N^{A'}\}$ and $b_i' \in \{I_N^{B'}\}$.
4. Decide for H_o, or H_1, depending on $\bar{z} < \Delta$, or $\bar{z} \geq \Delta$.

We tested the proposed algorithm on 16-*bit* stereo audio signals, sampled at Fs = 44.1KHz. The size of each patch (i.e., N) was fixed to 50 points, while the default values for (δ, θ) were set to (0.5, 10). Higher values for θ were also tested only for robustness evaluation, regardless of quality aspects. The state of the art proposes a framework for the evaluation of audio watermarking techniques (Gordy & Bruton, 2000). In this chapter, we referred to this framework and considered, in particular, two key factors: quality of the watermarked signal and robustness to mp3 compression. The evaluation of quality is an essential part in testing our strategy, since the basic idea was to guarantee the maximum rate of inaudibility of the patch. The tests were performed using a subjective score (i.e., a *MOS*) and the *SNR* of the watermarked signal vs. the host signal. The robustness of the proposed strategy was tested in two steps: at first, coding and decoding the watermarked signal with a commercial MP3 encoder at different rates (e.g., usually 128*Kbps*); secondly, trying the detection of the watermarked on the uncompressed signal. Quality and robustness cannot be evaluated separately. These factors are strongly correlated, that is, a decrease in quality causes an increase, in most cases significant, of robustness. All the performed tests showed good results. The idea of increasing the number of points of the patches reveals its successfulness. Good subjective quality appears since all the patches are below the audibility threshold for that signal ($SNR \leq 26$). At the same time, treating more points has the effect of shifting the patchwork-noise to low frequencies, where it has a lower probability to be filtered by the mp3 compression. Figure 17 shows different probability density functions (i.e., introduced as empirical probability density function (*Pdf*) in Bassia et al., 2001) of the random variable z, as described in the detection phase. The density function of z before the mp3 compression is compared with different behaviours (i.e., varying the dimension of the patch). This test shows clearly that higher dimensions of θ lead to lower alterations in the detection values. This results in a *Pdf* nearer to that of the uncompressed signal. We have also evaluated the error probability at different rates of compression (i.e., 128, 96 and 64 *Kbps*). Two kinds of errors can be individuated. The state of the art refering to them is terms of Type I (Rejection of H_o, when H_o is true) and Type II (Nonrejection of H_o, when H_1 is true) (Yeo et al., 2003). Type II errors seem to be the most impairing (the watermark is inserted (i.e., quality degradation) but the ownership cannot be proven). Figure 18 presents the Type II errors for a test audio signal. Clearly, the probability of rejection of H_o, when H_1 is true, decreases correspondently with the mp3-rate of compression.

The proposed strategy can be improved in two points: the modelling of the patch and the possibility of a wrong detection (that is, increasing the distance d, both the wrong detection and the inaudibility decrease). Actually, these step are quite coarse. Further studies have been centred on more refined mathematical mechanisms of patch shaping (i.e., such as the curve fitting) with respect to the minimum masking threshold. In the following, we deepen the aspects related to the second point and try a possible solution with an ad-hoc framework from the authors (i.e., named "Opportunity" strategy). The embedding steps are (1) choosing pseudorandomly two sets from Ψ; (2) defining a small quantity d; (3) adding d to a set and subtracting the same value from the other set. Specifically, $x_i^* = x_i + d, y_i^* = y_i + d$, where $x_i \in A$, $y_i \in B$ and $(A, B) \in \Psi$. The detection phase is based on a hypothesis testing, that is,

Figure 17. Probability density function of detection for the random variable z, varying the dimension of the patch with SNR = 26.

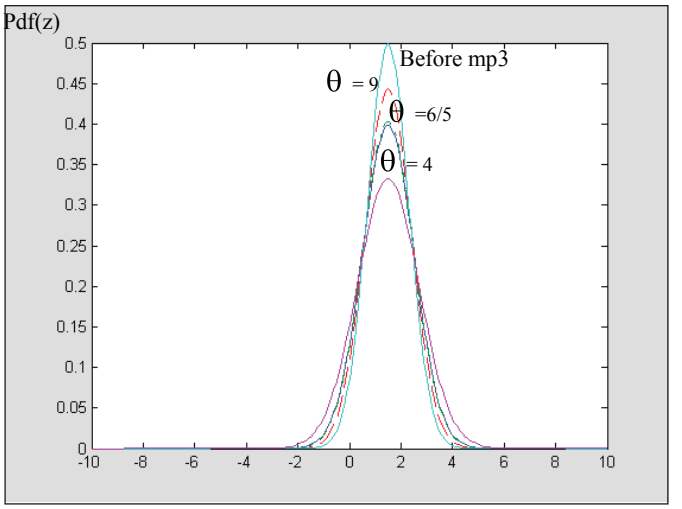


Figure 18. Error Probabilities for lossy compression at different rates

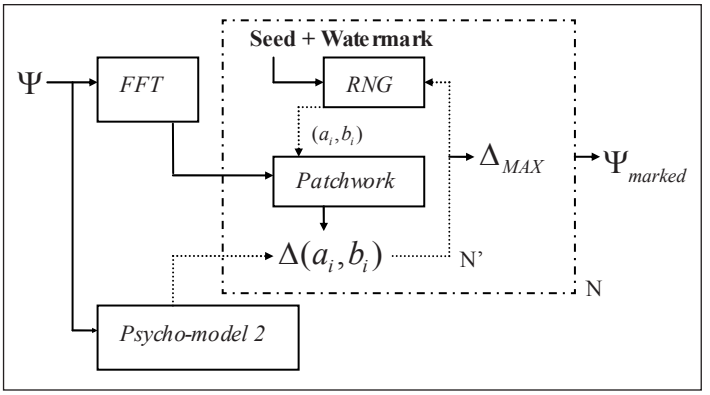
Type II Errors(%)	
MPEG I Layer III (128Kbps)	0.1
MPEG I Layer III (96Kbps)	0.7
MPEG I Layer III (64Kbps)	1.6

defined on a test statistic z and a test hypothesis H. The basic theory is as follows: let us suppose the *Pdf* of a random variable z (i.e., the test statistic) is a known function $f(Z, \theta)$ depending on the parameter θ. We wish to test the assumption $\theta = \theta_0$, against $\theta \neq \theta_0$. The assumption $\theta = \theta_0$ is denoted with the null hypothesis H_o, where $\theta \neq \theta_0$ is signified by the alternative hypothesis H_1. The purpose is to establish whether experimental evidence supports the rejection of H_o. Referring to the watermarking context: (1) the test statistic z is defined as a function of elements from A and B: $z = \vec{x} - \vec{y}$; (2) the distribution of z (e.g., $\Phi(z)$) is normal (Arnold, 2000), with an expected mean value $\overline{z} = 0$ in the unmarked and $\overline{z} = 2d$ in the marked case (i.e., in this case, $\Phi(z) = \Phi_m(z)$); (3) the test hypotheses (H_o, H_1) are defined and verified: H_o (i.e., the watermark is not embedded) and H_1 (i.e., the watermark is embedded and $\Phi(z) = \Phi_m(z)$). The main problem is the definition of d: the probability of a wrong decision remains, although the distribution in the watermarked case is shifted to the right (see Figure 15). A possible solution could be increasing the distance factor d, with the effect of having a lower probability of false rejection, but also a greater impairment on the audibility of the watermark in the host signal. The literature tries to solve this problem in several ways. In this sense, the idea that seems to work better is the spectral patchwork (Arnold, 2000; Kim, 2003; Yeo et al., 2003), with a dataset of $2N$ Fourier coefficients. The relationship between d and the elements of the dataset is multiplicative. The parameter d is adaptively chosen to prevent

perceptual audibility, based on the characteristics of the audio signal (i.e., it introduces for the first time the concept of power density function in the hypothesis tests). In Kim (2003), the patchwork algorithm is applied to the coarsest wavelet coefficients, providing a fast synchronization between the watermark embedding and detection. While in Yeo et al. (2003) the modified patchwork algorithm (MPA) is presented. Such an approach is very robust due to three attributes: factor d is evaluated adaptively and is based on sample mean and variance; the patch size in the transformed domain is very little: this guarantees good inaudibility; finally, a sign function is used to enhance the detection rate. It is stated that the watermark embedding can be considered as an operation of channel coding (Muntean et al., 2002): the watermark should adapt to the characteristics of the transmission channel (i.e., the host signal in which the watermark should be embedded). In case of audio contents, what is usually considered as an impairment, the sensibility of the human ear, can be used as a way to spread and dimension the watermark. Given a signal S, it is possible to recover its minimum masking threshold and use it to opportunely shape the watermark-noise. Several methods, outside the patchwork fashion, have been proposed that make use of psychoacoustic models to guarantee perceptual inaudibility of the mark. Opportunity refers and applies this approach in a patchwork schema, maximizing a cost function that represents how much the watermark impairs the original signal. The minimum masking threshold is recovered using the *psychoacoustic model 2* (ISO MPEG-AUDIO). The proposed strategy works as follows (see Figure 19): (1) the value of d is fixed *a priori* to minimize the possibility of false rejection; (2) given Ψ, its minimum masking threshold is recovered; (3) the secret key and the watermark are mapped to the seed of a random number generator (RNG) to find a couple (a_i, b_i); (4) the patchwork approach is applied to the DFT coefficients of the host signal corresponding to (a_i, b_i), using the distance factor d; (5) a cost function $\Delta = \Delta(a_i, b_i)$ if defined, that represents the global distance between the minimum masking threshold and the amplitude of the patch (i.e., the condition $\Delta > 0$ must be verified by default, otherwise a new set (a_i, b_i) is generated); (6) the process of generation of (a_i, b_i) (i.e., Steps (3)–5)) is iterated N' times to find the couple $(a_{i_\min}, b_{i_\min}) : \Delta = \Delta(a_{i_\min}, b_{i_\min}) = \Delta_{MAX}$; (7) the entire process is repeated to embed N couples of patches $(i = 1...N)$ and the watermarked signal Ψ_{marked} is obtained. The phase of detection is as follows: (1) define H_o (the watermark in not embedded) and H_1 (the watermark is embedded; (2) map the seed and the watermark to an *RNG* and generate the two global sets $I_N^{A'}$ and $I_N^{B'}$ (e.g., containing the elements $(a_i, b_i), i = 1...N$); (3), fix a threshold Δ for the detection, and evaluate the mean value (i.e., $\bar{\cdot} = E(\cdot)$) of the random variable $z = a_i' - b_i'$, for $a_i' \in \{I_N^{A'}\}$ and $b_i' \in \{I_N^{B'}\}$; (4) decide for H_o, or H_1, depending on $\bar{z} < \Delta$, or $\bar{z} \geq \Delta$.

We tested the proposed algorithm on 16-bit stereo audio signals, sampled at $Fs = 44.1KHz$. The size of the patch (i.e., N) was fixed to 100 couple of points while the default value for d was set to 0.25 (e.g., where the original algorithm (Bender et al., 1996) selected a fixed value 0.15 in the experiments). Higher values for d were also tested only for robustness evaluation, regardless of quality aspects, to test the rate of probability rejection. The performed investigations on the proposed strategy have the aim of showing *good audio quality* (e.g., inaudibility of the mark) of the watermarked signals and, at the same time, *low probability of false rejection* when the audio signal is subjected to postprocessing. In the following, the results of such tests are presented. The evaluation of quality is an essential part in testing our strategy, since the basic idea was to guarantee the maximum rate of inaudibility of the patch. Good subjective quality appears since all the patches are below the audibility threshold for that signal ($SNR \leq 26$). We also compared the marked/unmarked signals using the PEAQ strategy (ITU-R BS.1387-1): Opportunity shows a

Figure 19. "Opportunity" strategy



better performance with respect to the other tested strategies with default parameters (see Figure 20). The second aim of the tests was to verify a decreasing in the probability of false rejection. Two kinds of errors can be individuated. The state of the art refers to them in terms of Type I (Rejection of H_o, when H_o is true) and Type II (Nonrejection of H_o, when H_1 is true) (Yeo et al., 2003). Type II errors seem to be the most impairing (the watermark is inserted (i.e., quality degradation) but the ownership cannot be proven). Figure 21 presents the Type II error for a noncompressed test audio signal (test1.wav). We also prove the efficiency of Opportunity when compressing-decompressing (i.e., at different rates of compression (i.e., 128, 96 and 64 *Kbps*)) and resynchronizing the same test signal (see Figure 22).

Figure 20. PEQ—MOS rating

Opportunity	4.77
MPA [YK03]	4.00
Arnold [A00]	3.23

Figure 21. Error probabilities for non-compressed signal (test1.wav)

Type II Errors(%)	
Opportunity	0.1
MPA [YK03]	0.7
Arnold [A00]	1.6

Figure 22. Error probabilities for lossy compression at different rates (test1.wav)

Type II Errors(%) ***-Opportunity-***	
MPEG I Layer III (128Kbps)	0.32
MPEG I Layer III (96Kbps)	0.68
MPEG I Layer III (64Kbps)	0.81

CONCLUSION

This chapter has presented an overview of the use of masking models in watermarking techniques, which is a relatively novel and very interesting field of research. The aim of such techniques is to guarantee a high level of robustness and bandwidth of the watermark, as well as fidelity of the watermarked signal. The general scheme for audio content has been presented both in the embedding and the detection phase. The approach is very similar both for audio and image contents. To also give an idea on how to design, implement, and verify techniques based on such an approach, two techniques on audio *watermarking* are proposed

in the last section. Such strategies are both based on a two-set method, as presented in Bender et al. (1996) for image watermarking, but combined with a masking model. The development of new techniques jointly based on classic techniques and masking models seems to be a trend for the future. The strategies have been compared to others in terms of robustness and transparency of the watermark.

REFERENCES

Arnold, M. (2000). Audio watermarking: Features, applications and algorithms. *IEEE International Conf. Multimedia and Expo*, *2*, 1013-1016.

Arnold, M., & Schiltz, K. (2002). Quality evaluation of watermarked audio tracks. *SPIE Electronic Imaging*, *4675*, 91-101.

Bassia, Pitas, I., & Nicholaidis, N. (2001). Robust audio watermarking in the time domain. *IEEE Transaction on Multimedia*, *3*, 232-241.

Bender, W., Gruhl, D., Morimoto, N., & Lu, A. (1996). Techniques for data hiding. *IBM SYSTEMS JOURNAL*, *35*(3-4).

Boney, L., Twefik, A.H., & Hamdy, K. N. (1996). Digital watermarks for audio signal. *International Conference on Multimedia Computing and Systems*, 473-480.

Brandenburg, K., & Stoll, G. (1992). *The ISO/MPEG-audio codec: A generic standard for coding of high quality digital audio.* 92nd AES-Convention, preprint 3336.

Campbell, F. W., & Kulikowski, J. J. (1966). Orientation selectivity of the human visual system. *Journal of Physiology*, *187*, 437- 445.

Chen, B., & Wornell, G. W. (1999). Dither modulation: A new approach to digital watermarking and information embedding. *Proceedings of the SPIE –Security and watermarking of multimedia contents*, *3657*, 342-353.

Cox, I. J., Kilian, J., Leigton, F. T. & Shamoon, T. (1996). Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing, 6*, 1673-1687.

Craver, S., Memon, N., Yeo, B.-L., & Yeung, M. (1996). Can invisible watermarks resolve rightful ownerships? *IBM Research Technical Report RC 20509 IBM CyberJournal*.

Craver, S., Memon, N., Yeo, B.-L., & Yeung, M. (1997). Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks and implications. *IBM Research Technical Report RC 20755 IBM CyberJournal*.

Cvejic, N., Keskinarkaus, A., & Seppanen, T., (2001). Audio watermarking using m-sequences and temporal masking. *IEEE Workshops on Applications of Signal Processing to Audio and Acoustics*, 227-230.

Delaigle, J. F., Devleeschouwer, C., Macq, B., & Langendijk, L. (2002). Human visual system features enabling watermarking. *Proceedings of Multimedia and Expo*, ICME02, 489-492.

Eggers, J. J. Su, J. K., & Girod, B. (2000). A blind watermarking scheme based on structured codebooks. *Secure Images and Image Authentication-IEE Colloquium*, 4/1-4/6.

Ferwerda, J. A., Pattanaik, S. N., Shirley, P. & Greenberg, D. P. (1997). *Computer Graphics*, *31*, 143-152.

Gordy, J. D., & Bruton, L. T. (2000). Performance evaluation of digital audio watermarking algorithms. In *Proceedings of the 43rd Midwest Symposium on Circuits and Systems*.

Herkiloglu, K., Sener, S., & Gunsel, B. (2004). Robust audio watermarking by adaptive psychoacoustic masking. In *Proceedings of 12th IEEE*

Conference on Signal Processing and Communications Applications, 29-32.

Huang, J., & Shi, Y. Q. (2002). Adaptive image watermarking scheme based on visual masking. *Electronic Letters*, *34*(8), 748-750.

ISO/IEC Joint Technical Committee 1 Subcommittee 29 Working Group 11. (1993). *Information technology-coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, Part 3:Audio*. ISO/IEC 11172-3.

Kazakeviciute, G., Januskevicius, E., Rosenbaum, R., & Schumann, H. (2005). Tamper-proof image watermarking, based on existing public key infrastructure. *INFORMATICA*, 16(1).

Kim, H. (2000). Stochastic model based audio watermark and whitening filter for improved detection. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, *4*, 1971-1974.

Kim, H. J. (2003). *Audio watermarking techniques*. Pacific Rim workshop on Digital Steganography.

Kim, H. O., Lee, B. K., & Lee, N. Y. (2001). *Wavelet-based audio watermarking techniques: Robustness and fast synchronization*. Research Report 01-11, Division of Applied Mathematics-Kaist.

Kim, H. J., & Yeo, I. (2003). Modified patchwork algorithm: A novel audio watermarking scheme. *Transactions on Speech and Audio Processing*, *11*(4), 381-386.

Lambrecht, C., & Verscheure, O. (1996). Perceptual quality measure using a spatio-temporal model of the human visual system. *Proceedings of the SPIE*, *2668*, 450-461.

Legge, G. E. & Foley, J. M. (1980). Contrast masking in human vision. *Journal Optical Society*, *70*, 1458-1470.

Lubin, J. (1993). The use of psychophysical data and models in the analysis of display system performance. In A. B. Watson, *Digital images and human vision* (pp. 163-178).

Macq, B., & Quisquater, J. (1995). Cryptology for digital TV broadcasting. *Proceedings of the IEEE*, *83*, 944-957.

Mansour, M. F., & Tewfik, A. H. (2001). Audio watermarking by time-scale modification. *International Conference on Acoustics, Speech, and Signal Processing*, 3, 1353-1356.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.

Mitchell, D. S., Bin, Z., Tewfik, A. H., & Boney, L. (1998). Robust audio watermarking using perceptual masking, *IEEE Signal Processing*, *66*(3), 337-355.

Muntean, T., Grivel, E., Nafornita, I., & Najim, M. (2002). Audio digital watermarking for copyright protection. *International Workshop on Trends and Achievements in Information Technology*.

Pan, D. (1995). A tutorial on MPEG/audio compression. *IEEE MultiMedia*, *2*(2), 60-74.

Pantle, A., & Sekuler, R. W. (1969). Contrast response of human visual mechanisms sensitive to orientation and direction of motion. *Vision Res.*, *9*, 397-406.

Pennebaker, W. B., & Mitchell, J. L. (1993). *JPEG: Still image data compression standard*. Van Nostrand Reinhold.

Petitcolas, F. A. P., Anderson, R. J. & Kuhn, M. G. (1999). Information hiding—A survey. *Proceedings of the IEEE—special issue on protection of multimedia content*, *87*(7), 1062-1078.

Poirson, A., & Wandell, B. (1993). Appearance of colored patterns. *Journal of the Optical Society of America*, *10*(12), 2458-2470.

Probst, R., Lonsbury-Martin, B. L., & Martin, G. K. (1991). A review of otoacoustic emissions. *Journal of Acoustic Society*, *89*, 2027-2067.

Quan, X., & Zhang, H. (2004). Audio watermarking for copyright protection based on psychoacoustic model and adaptive wavelet packet decomposition. *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing*, 282-285.

Recommendation ITU-R BS.1387-1, 11/01.

Robert, A., & Picard, J. (2005). On the use of masking models for image and audio watermarking. *IEEE Transaction on Multimedia*, *7*(4), 727-739.

Robinson, D. J. M., & Hawksford, M. J. (1999). Time-domain auditory model for the assessment of high-quality coded audio. *107th Convention of the Audio Engineering Society* - preprint 5071.

Smith & Abel (1999). *Bark and ERB Bilinear Transforms*.

Swanson, M. D., Bin, Z., & Tewfik, A. H. (1998). Multiresolution scene-based video watermarking using perceptual models. *IEEE Journal on Selected Areas*, *16*(4), 540-550.

Tilki, J. F., & Beex, A. A. (1996). Encoding a hidden digital signature onto an audio signal using psychoacoustic masking. In *Proceedings 1996 7th International Conference on Signal Processing*. 476-480.

Tovée, M. J. (1996). *An introduction to the visual system*. Cambridge University Press.

Trichili, H., Bouhlel, M.-S., Solaiman, B., & Kamoun, L. (2003). Exploitation of the HVS features for the enhancement of image watermarking techniques. *ISPA03*, *2*, 1076-1081.

Wang, R. D., Xu, D. W., & Qian, L. (2005). Audio watermarking algorithm based on wavelet packet and psychoacoustic model. *PDCAT 2005*, 812-814.

Wu, M., & Liu, B. (1998). Watermarking for image authentication. *Proceedings of 1998 IEEE International Conference on Image Processing*, *2*, 437-441.

Wu, C. P., Su, P. C., & Kuo, C. C. J. (2000). Robust and efficient digital audio watermarking using audio content analysis. *Security and Watermarking of Multimedia Contents (SPIE)*, *3971*, 382-392.

Xie, G., & Shen, H. (2005). Toward improved wavelet-based watermarking using the pixel-wise masking model. *ICICP2005*, *1*, 689-692.

Xing, H., & Scordilis, M. S., (2005). Improved spread spectrum digital audio watermarking based on a modified perceptual entropy psychoacoustic model. *Proceedings of IEEE Southeast Conference*, 283-286.

Yeo, I. K., & Kim, H. J. (2003). Modified patchwork algorithm: A novel audio watermarking scheme. *IEEE Transaction On Speech and Audio Processing*, *11*(4).

Zwicker, E., & Zwicker, U. T. (1991). Audio engineering and psychoacoustics: matching signals to the final receiver: The human auditory system. *Journal of Audio Engineering Society*, *39*(3), 115-126.

ENDNOTES

1 The expression $E(o)$ indicates the operation of mean value in statistics.

2 Including components below the absolute hearing threshold, or components too close to others to be heard.

3 This step involves the human auditory system.