An Audio Patchwork Shaping Framework with Psychoacoustic Model 2

Maurizio Murroni[†]*, Member, IEEE, Mirko Luca Lobina*[†]*, Student Member, IEEE, and Ivana Schirru*[†]

[†] DIEE (Department of Electric and Electronic Engineering) – University of Cagliari

ABSTRACT

In this paper we propose an adaptive watermarking framework for audio contents, based on a patchwork approach. The patchwork strategy was originally described in [1] 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 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 versus the minimum masking threshold, obtained with the *psychoacoustic model 2*. For this purpose, the proposed strategy works on the Fourier coefficients. The robustness to lossy compression is satisfied, considering multipoint patches [1]. The proposed technique does not need the original signal for the detection of the watermark. Experiments and simulation results verify the effectiveness of our approach.

1. INTRODUCTION

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 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 [2]. Two characters are mostly 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. Secondly the watermark must be robust against intentional or unintentional attacks. One of the most impairing attack is the signal processing, and specifically the lossy compression. Such a compression guarantees enhanced portability of digital information, but can have an undesirable effect on the embedded watermark. The developing of our strategy is accomplished referring constantly to these two features.

In this paper we refer to an adaptive approach of patchwork algorithm. The patchwork is originally presented in [1] 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 ([2]-[5]).

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 [1]: treating patches of several points have the effect of shifting the noise to low frequencies, where there is lower probability to be filtered by lossy compression techniques. The right dimension is fixed by comparing the spectrum of watermarked signal to the minimum masking threshold, as obtained referring to the psychoacoustic model 2 [6]. 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 art, referring to the framework proposed in [7]. The patchwork shaping framework shows particular good results in terms of robustness to compression and quality. The paper is organized as follows. Section 2 presents the state of art of adaptive patchwork algorithms. Section 3 introduces the watermark shaping despite of the threshold of audibility. Section 4 illustrates our technique. Section 5 presents tests and results, while in Section 6 the conclusions are drawn.

2. ADAPTIVE PATCHWORK APPROACH

The patchwork strategy is a two-set method, that is it makes different two sets from a host signal [4]. This difference is used to verify, or not, a hypothesis H_o (e.g., the watermark is embedded).

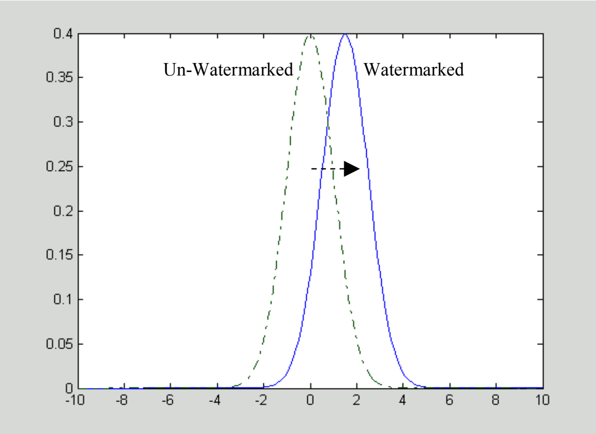

Figure 1: Distribution of the mean difference of the samples in Un-Watermarked and Watermarked signals.

The original strategy [1] 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 [2]. In particular the distribution of the sample value is assumed as normal (See *Fig.1*). Recent approaches modify the original strategy to better take into account the human ear sensibility to noise interferences. These methods can be classified in temporal and spectral approaches, depending on the domain where the watermark is embedded. In [5] 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)*. [2] – [3] – [4] propose spectral patchwork approaches. In particular, [2] works 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, basing on the characteristics of the audio signal (i.e., it is introduced for the first time the concept of power density function in the hypothesis tests).

In [3] the patchwork algorithm is applied to the coarsest wavelet coefficients, providing a fast synchronization between the watermark embedding and detection. While, in [4] the Modified Patchwork Algorithm (i.e., MPA) is presented. Such 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 set 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 [3].

3. WATERMARK SHAPING AND THE PSYCHOACOUSTIC MODEL

The association between a watermarking algorithm and a noisy communication system is not new [8]. 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, what 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 sensible 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 (i.e., ATH). Such curve (i.e., referred to as *quiet curve* [9]) 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 [10].

Several methods, outside the patchwork fashion, have been proposed that make use of *psychoacoustic models* to guarantee perceptual inaudibility of the mark [11], [9], [12]. Usually, such methods shape the watermark, referring to the quiet curve. The filtered watermark signal is scaled in order to embed the watermark noise below the quiet curve [13]. In addition, it is possible to increase the noise energy of the watermark, referring undeniably to the minimum threshold of audibility. Such 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

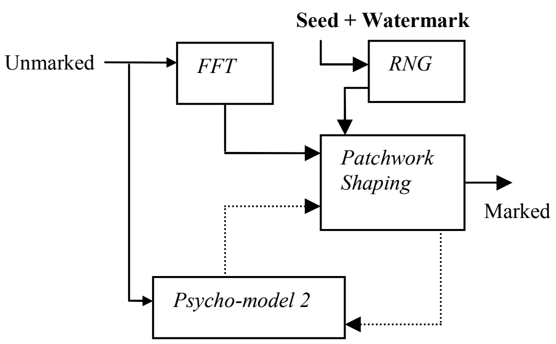

Figure 3: Steps (1-4) of the Patchwork shaping algorithm.

any of the layers of compression. However, only *model 2* includes specific modifications to accommodate Layer III. In this paper, we refer to the *model 2*, differently from the past approaches.

4. PROPOSED STRATEGY

As already stated, the proposed patchwork strategy modifies two set of *N* elements/coefficients from the original signal (i.e., $signal_{Un\text{-}Marked}$). The $signal_{Marked}$ strongly belongs to the correspondent $signal_{Un\text{-}Marked}$. 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 work 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 *Fig.3*) can be summarized as follows:

1. Evaluate the threshold of minimum audibility for the $signal_{Un\text{-}Marked}$, 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 $signal_{Un\text{-}Marked}$, corresponding to the index sets I_N^A and I_N^B.
4. The original amplitude of the patch and the number of re-touched 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 *Fig.3*).
5. The time-domain representation of the output signal is found, applying an Inverse DFT to the $signal_{Marked}$.

The phase of detection is as follows:

1. Define two test hypothesis: H_o (the watermark in not embedded) and H_1 (the watermark is embedded).
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., $\bar{z} = E(\cdot)$) 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 $\bar{z} \geq \Delta$.

5. TESTS AND RESULTS

We tested the proposed algorithm on 16-bit stereo audio signals, sampled at Fs = 44.1*KHz*. The size of the 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 art proposes a framework for the evaluation of audio watermarking techniques [7]. In this work, 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 versus 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 can not 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 ($SNR \leq 26$).

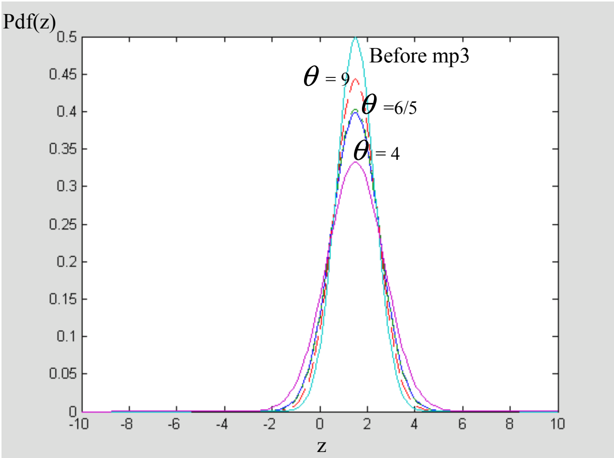

Figure 4: Probability density function of detection for the random variable *z*, varying the dimension of the patch with SNR = 26.

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 4 shows different probability density functions (i.e., introduced as *empirical Pdf* in [5]) 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 art refers to them is terms of Type I (Rejection of H_o, when H_o is true) and Type II (Non-Rejection of H_o, when H_1 is true) [4]. Type II errors seem to be the most impairing (the watermark is inserted (i.e., quality degradation) but the ownership can not be proven). Table I 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.

6. CONCLUSIONS

In this paper an audio watermarking framework has been presented, that is based on a patchwork approach. The core of the proposed technique is the shaping of the patchwork, performed referring to psychoacoustic *model 2*. This results in higher robustness and inaudibility of the patch noise. The strategy was evaluated in terms of robustness to lossy compression, quality. Good results were obtained during the tests with 44.1*Khz* audio and speech traces. The proposed strategy can be improved in the modelling of the patch.

Table I: 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

Actually this step is quite coarse. Further studies will be centred on more refined mathematical mechanisms of patch-shaping (i.e., such as *the curve fitting*) respect to the minimum masking threshold.

7. REFERENCE

[1] W.Bender, D.Gruhl, N.Morimoto, and A.Lu, “Techniques for data hiding,” *IBM System Journal*, Vol.35, No.3-4, pp.313-335, 1996.

[2] M.Arnold, “Audio watermarking: features, applications and algorithms,” *IEEE Int. Conf. Multimedia and Expo 2000*, Vol.2, pp.1013-1016, 2000.

[3] Hong Oh Kim, Bae Keun Lee, and Nam Yong Lee, "Wavelet-based audio watermarking techniques: robustness and fast synchronization,” *Research Report 01-11*, Division of Applied Mathematics-Kaist.

[4] In-Kwon Yeo and Hyoung Joong Kim, “Modified patchwork algorithm: A novel audio watermarking scheme, “ IEEE Trans. On Speech and Audio Processing, Vol.11, No.4, 07/2003.

[5] P.Bassia, I.Pitas, and N.Nicholaidis, “Robust audio watermarking in the time domain,” *IEEE Transaction on Multimedia*, Vol.3, pp.232-241, 06/2001.

[6] ISO/IEC Joint Technical Committee 1 Subcommittee 29 Working Group 11, *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, 1993.

[7] J. D. Gordy and L. T. Bruton, “Performance Evaluation of Digital Audio Watermarking Algorithms,” *Proceedings of the 43rd Midwest Symposium on Circuits and Systems*, Lansing MI, USA, Aug. 2000;

[8] T.Muntean, E.Grivel, I.Nafornita, and M.Najim, “Audio digital watermarking for copyright protection,” *International Workshop on “Trends and Achievements in Information Technology”*, 05/2002.

[9] N.Cevjic, A.Keskinarkaus, and T.Seppanen, “Audio watermarking using m-sequences and temporal masking,” *IEEE Workshops on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, pp.227-230, 2001.

[10] E.Zwicker and U.T.Zwicker "Audio Engineering and psychoacoustics: matching signals to the final receiver: the Human Auditory System," *Journal of Audio Engineering Society*, Vol.39, No.3, pp.115-126, 03/1991.

[11] L.Boney, A.H.Twefik, and K.N.Hamdy, “Digital watermarks for audio signal,” *International Conference on Multimedia Computing and Systems*, Hiroshima, Japan, pp.473-480, 1996.

[12] M.Arnold and K.Schiltz, “Quality evaluation of watermarked audio tracks,” *SPIE Electronic Imaging*, Vol.4675, pp.91-101, 2002.

[13] Hyoung Joong Kim, “Audio watermarking techniques,” *Pacific Rim workshop on Digital Steganography*, Kitakyushu, 07/2003.

[14] D. Pan, "A tutorial on MPEG/audio compression," *IEEE MultiMedia*, Vol.2, No.2, pp. 60--74, 1995.