# Perceptual audio watermarking driven by Human Auditory System

Federica Battisti*, Marco Carli* and Claudia Rinaldi†
*Engineering Department, Università degli Studi Roma TRE, Roma, Italy
Email: federica.battisti,marco.carli@uniroma3.it
†Center of Excellence DEWS, University of L'Aquila, L'Aquila, Italy
Email: claudia.rinaldi@univaq.it

***Abstract*— The aim of this work is the design of a novel audio watermarking technique based on the Linear Predictive Coding approach and on the psychoacoustic model defined in the MPEG-I standard. The embedding imperceptibility is achieved by exploiting the frequency masking effect. The evaluation of the perceived audio quality is performed by means of the standardized ITU-R BS.1387 quality assessment method. The effectiveness of the proposed method is tested in presence of attacks.**

## I. INTRODUCTION

The problem of copyright protection of digital media has attracted the interest of the scientific and business communities. Among others, a promising solution to deal with this problem is the watermarking process. The data to be protected is modified (or *marked*) with ownership information in a, usually, imperceptible way. Depending on whether it requires access to the original signal and/or the actual watermark, the detection process may be described as informed, semi-blind, or blind. Most of existing watermarking techniques have been designed for image and video signals [1]–[4]. Recent works are based on the exploitation of human perception features for optimizing the embedding performances [5], [6] also in terms of annoyance of embedding artifacts.

Compared to hiding information into 2D or 3D data, audio watermarking is more challenging due to the high sensitivity of the human auditory system to audio signal modifications. To this aim, many audio watermarking systems try to exploit the outcomes of the studies on the Human Auditory System (HAS); an extensive overview of the human acoustic perception can be found in [7]. Many contributions have been presented on this topic exploiting different techniques. Spread spectrum [8], phase coding [9], temporal masking [10], patchwork [11], and echo hiding [12]. Here we present novel audio watermarking techniques based on the use of Linear Predictive Coding (LPC) and on the psychoacoustic model defined in the MPEG-I standard. The embedding imperceptibility is achieved by exploiting the frequency masking effect.The evaluation of the perceived audio quality is performed by means of the standardized ITU-R BS.1387 assessment method.

## II. PROPOSED METHOD

In this paper we propose two watermarking techniques: the first one exploits the perceptual characteristics of the Human Auditory System (HAS) while the second, based on the previous one, includes the Linear Predictive Codes (LPC) for increasing the watermarking effectiveness. The two approaches are described in the following subsections.

### A. *Perceptual watermarking*

The algorithm can be summarized as follows:

1) The original audio signal $s(n)$ is partitioned in blocks of $384$ samples.
2) The watermark sequence, $b(n)$, of length 6 bits, undergoes a spreading procedure with a pseudo-random sequence $r(n)$ of length 12 bits, resulting in a watermarked spread sequence $w(n)$ of length 72 bits.
3) $s(n)$ is processed according the psychoacoustic model used in MPEG-I [13] for extracting the Signal to Mask Ratio (SMR).
4) $w(n)$ is multiplied by the SMR for adapting it to the perceptual characteristics thus obtaining $W(n)$.
5) The processed watermark is added to Fourier coefficients corresponding to negative values of SMR as:
$$f_i' = f_i + \alpha * W(n) \quad (1)$$
where $f_i'$ is the watermarked Fourier coefficient, $f_i$ the original Fourier coefficient and $\alpha$ is the watermark strength.
6) The Inverse FFT is computed and the watermarked audio signal $s_w(n)$ is reconstructed.

The detection procedure is base on the following steps:

1) The watermarked audio signal $s_w(n)$ is partitioned in blocks of $384$ samples.
2) The FFT is computed for each block.
3) The correlation between the watermarked Fourier coefficients and the sequence $r(n)$ used for spreading of the message $b(n)$ is computed and compared to a threshold $\tau$ for extracting the watermark bits $b_k$:
$$b_k = \begin{cases} 0, & c \leq \tau \\ 1, & c > \tau \end{cases} \quad (2)$$
4) The procedure is iteratively repeated for extracting the entire message.

978-1-4673-6143-9/13/$31.00 ©2013 IEEE

### B. LPC-based watermarking scheme

In this second approach LPC is used to strengthen the watermarking algorithm. The residual is used for hosting the watermark as shown in Figure 1. The LPC method is based

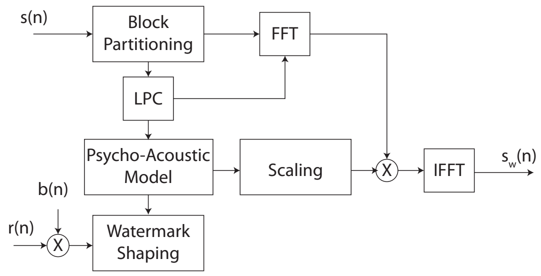



Fig. 1. LPC watermarking scheme.

on the approximation of the original audio signal $x(n)$ as a linear combination of $p$ past audio samples:

$$x(n) = \sum_{i=1}^{p} a_i(n-i) + e(n) \tag{3}$$

where the coefficients $a_1, a_2, .., a_n$ are assumed to be constant during the analysis of the audio segment and $e(n)$ is an excitation or the residual signal of $x(n)$. The watermarked audio signal $\widetilde{x}(n)$ is:

$$\widetilde{x}(n) = \sum_{i=1}^{p} a_i(n-i) + \ddot{e}(n) \tag{4}$$

where $\ddot{e}(n)$ is the watermarked residual. The detection of the watermark is performed by extracting the watermark from the residue obtained by LPC encoding of the watermarked audio signal.

## III. Experimental results

To test the effectiveness of the proposed method, a set of three audio files with different characteristics have been selected: *REM* (only instrumental sounds), *Speech* (only voice), and *Sugar* (both instrumental and vocal sounds). In Figure 2, the Power Spectral Density (PSD) of the original signal and of the signal watermarked with additive and perceptual techniques are shown.

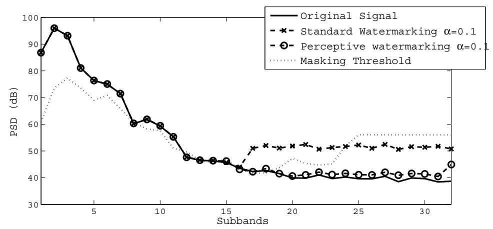
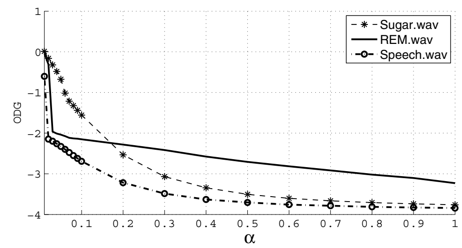
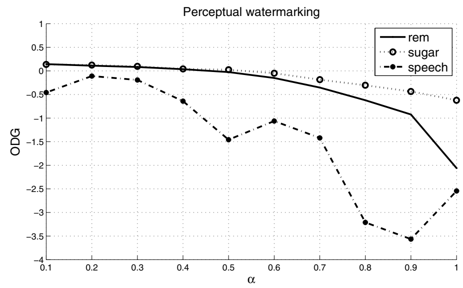



Fig. 2. Power Spectral Density comparison between perceptive watermarking and simple watermarking.

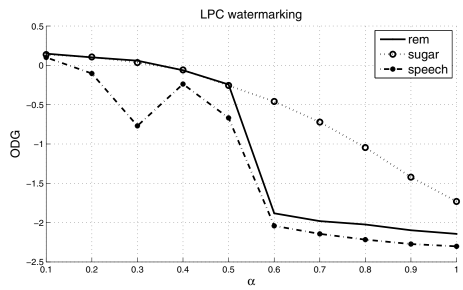
Figure 3 shows the comparison between the Objective Difference Grade (ODG) evaluated on the three audio files for additive, perceptual and LPC watermarking with prediction order $p = 30$. ODG is evaluated on the basis of PEAQ (Perceptual Evaluation of Audio Quality) algorithm [14]. ODG assumes vales in the range $[-4, 0]$ where $-4$ indicates the lowest quality and 0 the highest one. As expected, perceptual watermarking shows better performances in terms of perceived audio quality. On the other hand, LPC watermarking, although bringing advantages with respect to additive watermarking, leads to a decrease in ODG with respect to perceptual watermarking. It is also possible to notice that the sound quality decreases when the watermark strength $\alpha$ increases and this trend is more evident in the case of perceptive and LPC watermarking than in the additive one.




(a) ODG for Additive watermarking.




(b) ODG for Perceptual watermarking.




(c) ODG for LPC watermarking.

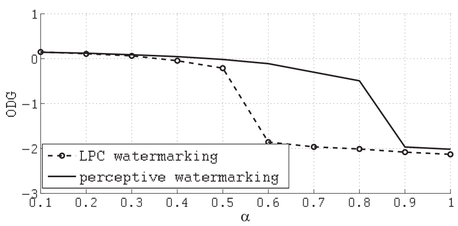
Fig. 3. Comparison between ODGs for the analyzed watermarking techniques.

The robustness of the proposed method has also been tested in terms of decoding Bit Error Probability (dBER), that is defined as, [15]:

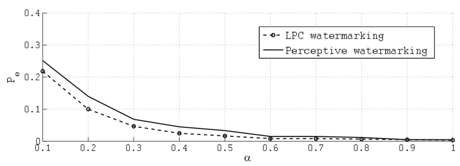
$$\text{dBER} = 1 - \frac{N_C}{N_E} \tag{5}$$

where $N_C$ is the number of correctly detected bits, while $N_E$ is the number of embedded bits.

For sake of clarity, Figure 4 presents the ODG and the dBER evaluated for one audio file (REM.wav) in order to better understand the differences between perceptual and LPC watermarking. Figure 4(a) shows that LPC watermarking is comparable to perceptual watermarking in terms of ODG only up to $\alpha = 0.5$, while for higher values of $\alpha$ the ODG for LPC watermarking decreases with respect to the same parameter evaluated for perceptual watermarking. Figure 4(b) shows the trend of the dBER parameter and, as expected, the robustness of both perceptual and LPC watermarking increases proportionally to the increase of the watermark strength $\alpha$. For this parameter, the LPC watermarking performs better for almost all values of $\alpha$.




(a) ODG for LPC watermarking of REM.wav.



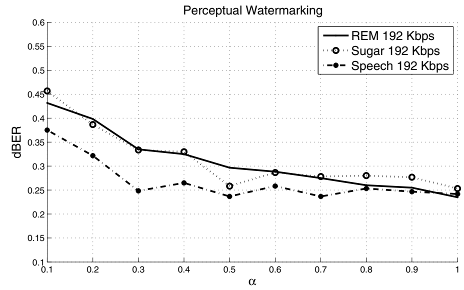

(b) $P_e$ of REM.wav for LPC.

Fig. 4. Comparison between the performarces of LPC watermarking vs perceptual watermarking.

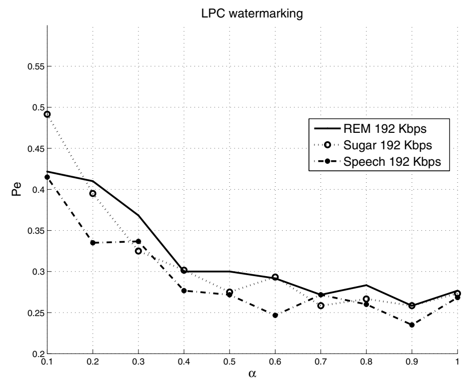
The proposed methods have also been tested under several attacks. Those were selected from the pool of available attacks in the StirMark Benchmark [16] that includes musical effects frequently used in pre-processing transmission or post-processing operation. In particular, after extensive experiments, we selected low pass filtering (15KHz), resample, and mp3 compression. It is important to notice that resampling and low-pass filtering cause a substantial degradation of the perceived audio quality. In this case, even if the attack is able to remove the watermark, the resulting audio file results to be of very low quality and consequently impossible to use.

Results obtained when the mp3 attack is performed are presented in Figure 5, 6. It is possible to notice that for high compression rates, perceptual and LPC watermarking schemes present similar performances, while the LPC watermarking scheme outperforms the MPEG one for lower compression rates and for $\alpha$ values larger than $0.3$.

Figure 8 reports the ODG and error probability for LPC watermarking method when the compression attack is performed.




(a) dBER for MPEG watermarking for mp3 compression to 192 Kbps.




(b) dBER for LPC watermarking for mp3 compression to 192 Kbps.

Fig. 5. dBER of LPC vs MPEG watermarking under mp3 attack to 192 Kbps.

## IV. Conclusions

In this paper we proposed the introduction of LPC into perceptual MPEG-1 watermarking based on the spread spectrum approach. Performances have been extensively investigated and the achieved results are encouraging in terms of perceived audio and decoding probability. Future works envisage the introduction of Independent Component Analysis for increasing the robustness of the algorithm.

## V. Acknowledgment

This work has been partially supported the EU FP7 Marie Curie IAPP project DIVeFor 'Digital Image and Video Forensics' (Grant Agreement No. 251677)

## References

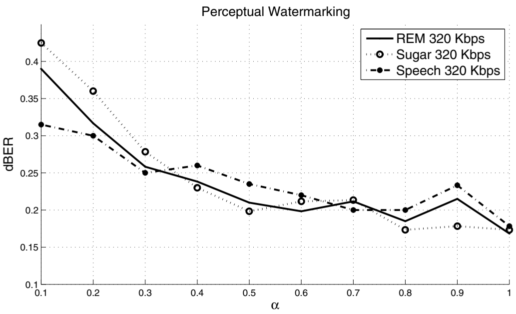
[1] A. Tewfik and M. Swanson, "Data hiding for multimedia personalization, interaction, and protection," *Signal Processing Magazine, IEEE*, vol. 14, no. 4, pp. 41–44, 1997.

[2] G. Boato, M. Carli, F. Battisti, M. Azzoni, and K. Egiazarian, "Difference expansion and prediction for high bit-rate reversible data hiding," *Journal of Electronic Imaging*, vol. 21, no. 3, pp. 033 013–1–033 013–9, 2012.

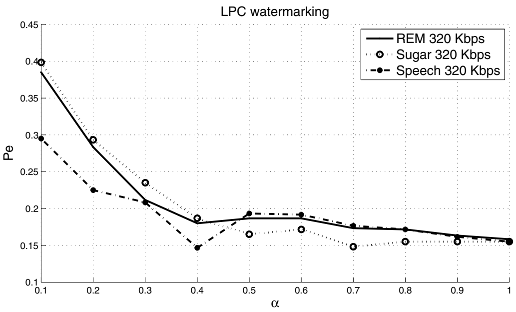
[3] F. Battisti, M. Cancellaro, M. Carli, G. Boato, and A. Neri, "Watermarking and encryption of color images in the fibonacci domain," in *SPIE Int. Conf. on Electronic Imaging, Image Processing: Algorithms and Systems VII*, 2008.

[4] I.J.Cox, G. Döerr, and T.Furon, "Watermarking is not cryptography," *Int. Work. on Digital Watermarking*, 2006.

[5] Q.Li and I.J.Cox, "Using perceptual models to improve fidelity and provide resistance to valumetric scaling for quantization index modulation watermarking," *IEEE Trans. on Information Forensics and Security*, vol. 2, no. 2, 2007.




(a) dBER for MPEG watermarking for mp3 compression to 320 Kbps.




(b) dBER for LPC watermarking for mp3 compression to 320 Kbps.

Fig. 6. dBER of LPC vs MPEG watermarking under mp3 attack to 320 Kbps.


[6] S. Maranò, F. Battisti, A. Vaccari, G. Boato, and M. Carli, "Perceptual data hiding exploiting between-coeffcient contrast masking," 2008.

[7] T. Painter and A. Spanias, "Perceptual coding of digital audio," in *Proceedings of the IEEE*, vol. 88, no. 4, Apr. 2000, pp. 451 –515.

[8] D. Kirovski and H. Malvar, "Robust spread-spectrum audio watermarking," in *IEEE Int. Conf on Acoustics, Speech, and Signal Processing*, 2001.

[9] L. Wang, S. Emmanuel, and M. Kankanhalli, "Emd and psychoacoustic model based watermarking for audio," in *IEEE Int. Conf. Multimedia and Expo*, 2010.

[10] Y. Xiang, D. Peng, I. Natgunanathan, and W. Zhou, "Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo-based audio watermarking," *Multimedia, IEEE Transactions on*, vol. 13, no. 1, pp. 2 –13, feb. 2011.

[11] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Trans. on Multimedia*, vol. 3, no. 2, pp. 232 –241, jun 2001.

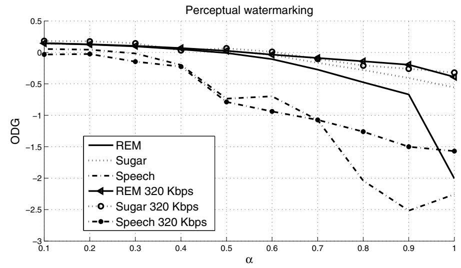
[12] O.-C. Chen and W.-C. Wu, "Highly robust, secure, and perceptual-quality echo hiding scheme," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 629 –638, 2008.

[13] MPEG, "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*, 1993.

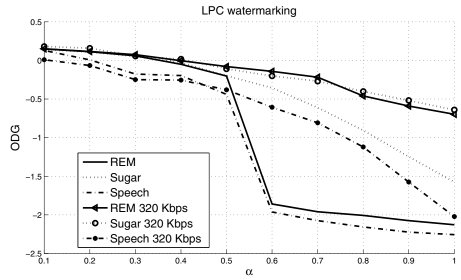
[14] P. Kabal, "An examination and interpretation of itu-r bs.1387: Perceptual evaluation of audio quality," 2003.

[15] H. Malik, A. Khokhar, and R. Ansari, "Improved watermark detection for spread-spectrum based watermarking using independent component analysis," in *ACM Work. on Digital Rights Management*, 2005.

[16] M. Steinebach, F. Petitcolas, F. Raynal, J. Dittmann, C. Fontaine, S. Seibel, N. Fates, and L. Ferri, "Stirmark benchmark: audio watermarking attacks," in *Int. Conf. on Information Technology: Coding and Computing*, 2001.

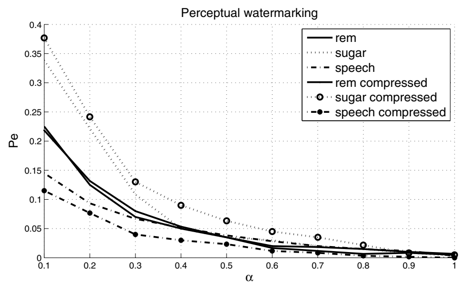




(a) ODG for MPEG watermarking for mp3 compression to 320 Kbps.

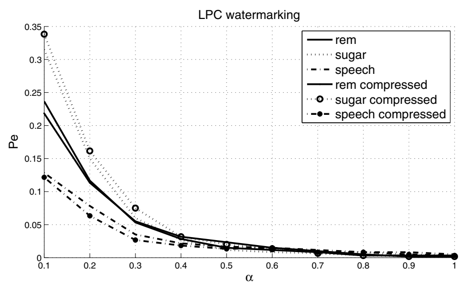



(b) ODG for LPC watermarking for mp3 compression to 320 Kbps.

Fig. 7. ODG of LPC vs MPEG watermarking under mp3 attack to 320 Kbps.




(a) dBER for MPEG alone watermarking with and without compressor attack..




(b) dBER for LPC watermarking with and without compressor attack.

Fig. 8. Comparison between dBER of the two watermarking schemes with and without compressor attack.