14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

DIGITAL AUDIO WATERMARKING FOR QoS ASSESSMENT OF MP3 MUSIC SIGNALS

F. Benedetto, G. Giunta, A. Neri, C. Belardinelli

Digital Signal Processing and Multimedia Communications Lab.,
Dept. of Applied Electronics, University of ROMA TRE,
Via della Vasca Navale 84, 00146 Rome, Italy, fax: +39 0655177026
e-mails: fbenedetto@ieee.org, giunta@ieee.org, neri@uniroma3.it, carlobelardinelli@yahoo.it

ABSTRACT

Digital watermarking of multimedia contents has become a very active research area over the last several years. In this work, we propose an audio watermarking signal processing technique to provide a quality assessment of the received audio signal after a coding/transmission process. Specifically, a fragile watermark is hidden in MP3-like host data audio transport stream (MPEG-1 layer III) using a spread-spectrum approach. At the receiving side, the watermark is extracted and compared to its original counterpart. Since the alterations endured by the watermark are likely to be suffered by the MP3-file, as they follow the same communication link (including coder and transport connection), the watermark's degradation can be used to estimate the overall alterations endured by the entire MP3-data. The Quality of Service assessment is based on the evaluation of the mean-square-error between the estimated and the actual watermark. The proposed technique has been designed for application to mobile multimedia communication systems. The results obtained through our simulation trials confirm the validity of such approach.

1. INTRODUCTION

In the past decade there has been an explosion in the use and distribution of digital multimedia data. PCs with Internet connections have made the distribution of multimedia data and applications much easier and faster [1]. In a networked environment like the World Wide Web, the crucial issue to be satisfied is the necessity to answer the ever-growing need to protect the intellectual property (copyright) of digital still images, video sequences, and audio from piracy attacks [2]. Therefore, there is an increase in concern over copyright protection of digital contents [3]. Digital media is more convenient than analog media in its usage, exchange of data, editing and storage. Moreover, digital media is *complete* without any distortion of data during the process of wide distribution through electric wave and network as well [4]. However, because of such characteristics, it is possible to duplicate the data and thus, many illegal trades often occur through the network. The aim of a controlled distribution of multimedia data can be reached developing suited signal processing techniques, such as digital watermarking. Digital watermarking of multimedia content has become a very active research area over the last several years [2].

In this contribution, we propose an *unconventional* use of audio watermarking procedure in order to provide a quality assessment of the received audio signal after a coding/transmission process. Although copyright protection was the very first application of watermarking, different uses have been recently proposed in the literature. Fingerprinting, broadcast monitoring, data authentication, multimedia indexing, content-based retrieval applications [5]-[12], are only a few of the new applications where watermarking can be usefully employed. When these techniques are used to preserve the copyright ownership with the purpose of avoiding unauthorized data duplications, the embedded watermark should be detectable. In particular, fragile watermarking has been recently proposed to blind estimate the Quality of Service (QoS) of a video-communication [7]. Most of the research on audio watermarking has been focused on either direct watermarking of the audio signal or bit stream embedding where the audio is represented in a compressed format [13]-[19]. When considering a watermarking scheme, depending on its specific application, different requirements need to be achieved. Many of the requirements for audio watermarking are similar to image watermarking, such as imperceptibility (inaudibility), robustness to signal alterations such as compression, filtering, and A/D and D/A conversion. This means that the perceptual invisibility of the superimposed mark onto the host data must be guaranteed. The choice of the embedding process should not affect the robustness of the embedded marks against malicious attacks. The proposed audio watermarking technique is here adopted to provide a quality assessment of the received audio signal after a coding/transmission process, without affecting the quality of the communication. Specifically, a fragile watermark is hidden in an MP3-like host data audio transport stream (i.e. MPEG-1 layer III coder [20]) using a spread-spectrum approach. At the receiving side, the watermark is estimated and compared with the original one. Since the alterations endured by the watermark are *likely* to be suffered also by the MP3-file, as they follow the same communication link, the watermark degradation can be used to estimate the overall alterations endured by the entire MP3-data. The proposed watermarking technique has been designed for application to mobile multimedia communication systems. Such a QoS index can be usefully employed for a number of different purposes in multimedia communications such as: control feedback to the sending user on the effective quality of the link; detailed information to the operator for billing purposes, etc...

The remainder of this work is organized as follows. Section 2 describes the signal processing algorithm used for the watermark embedding process, while Section 3 shows the quality evaluation setup, detailing the watermark estimation and introducing the employed metric for the quality-assessment procedure. Simulation results are finally provided in Section 4, before paper's conclusions drawn in Section 5.

2. AUDIO WATERMARKING SIGNAL PROCESSING FOR QoS ASSESSMENT

The principle scheme of the proposed audio watermarking procedure for quality assessment is reported in Fig. 1. The watermark embedding is performed in the temporal domain by means of a spread spectrum approach [10]. Specifically, a fragile watermark is hidden in an MP3-like host data audio transport stream (MPEG-1 layer III coder [20]) using a spread-spectrum approach. More in details, the original audio signal is divided into N blocks (audio-blocks) composed each by M samples and corresponding to two seconds of 44.1 KHz stereo audio signal. The watermark $w[k]$ that has been employed is a binary sequence consisting of a number of k bits. A set of uncorrelated pseudo-random noise (PN) vectors (one per each block and known to the receiver) is multiplied by the reference watermark (one for all the transmission session and known to the receiver) as follows:

$$w_i^{(s)}[k] = w[k] \cdot p_i[k], \quad i = 1, 2, ..., N \tag{1}$$

where $p_i[k]$ is the i-th PN vector and $w_i^{(s)}[k]$ is the spread version of the mark to be embedded in the i-th block. Like in spread-spectrum techniques, the use of different spreading PN vectors assures that the embedding of the mark is different from an MP3 block to the following one, so that the watermark can be considered as a perceptually inaudible modification of the audio signal. Moreover, the method is robust against permanent bit errors, due to either the physical network or its management (e.g. multi-path of the transmission channel, multi-user interference, excess loading factors, etc.). Let now be $G_i[k]$ = DCT$\{g_i[k]\}$ the DCT of the i-th audio block $g_i[k]$. The watermark, randomized by the PN vectors, is then multiplied by a scaling coefficient α (i.e. the power of the mark), and finally added to the DCT of each audio-block in the middle-high frequency region F, according to the following rule:

$$G_i^{(w)}[k] = \begin{cases} G_i[k] + \alpha w_i^{(s)}[k], & k \in F \\ G_i[k], & k \notin F \end{cases} \tag{2}$$

where $G_i^{(w)}[k]$ is the i-th watermarked audio block while α, the *power* of the mark, is a scaling factor that determines the watermark strength. By increasing its value, the mark becomes more audible, and a degradation of the original audio signal occurs. On the contrary, by diminishing its value, the mark can be easily removed by the coder and/or channel's errors. After the inverse DCT, the audio signal is coded by an MP3 coder (MPEG-1 Layer III coder at different compression ratios) and transmitted (see Fig. 1).

At the receiving side, the audio signal is first decoded; then, from the DCT of each watermarked audio block, a matched filter extracts the spread watermark, which is finally despread using the known PN vectors. After having extracted the received watermark, it is matched to the reference one, which is known at the receiving side, and the mean-square error (MSE), between the original mark and the received one, is used as an index of the degradation affecting the received watermark. In this way, the watermark is affected only by the channel's errors and at the receiver side the estimation of the degradations affecting the received mark can be used to provide a quality assessment of the received audio signal after a coding/transmission process.

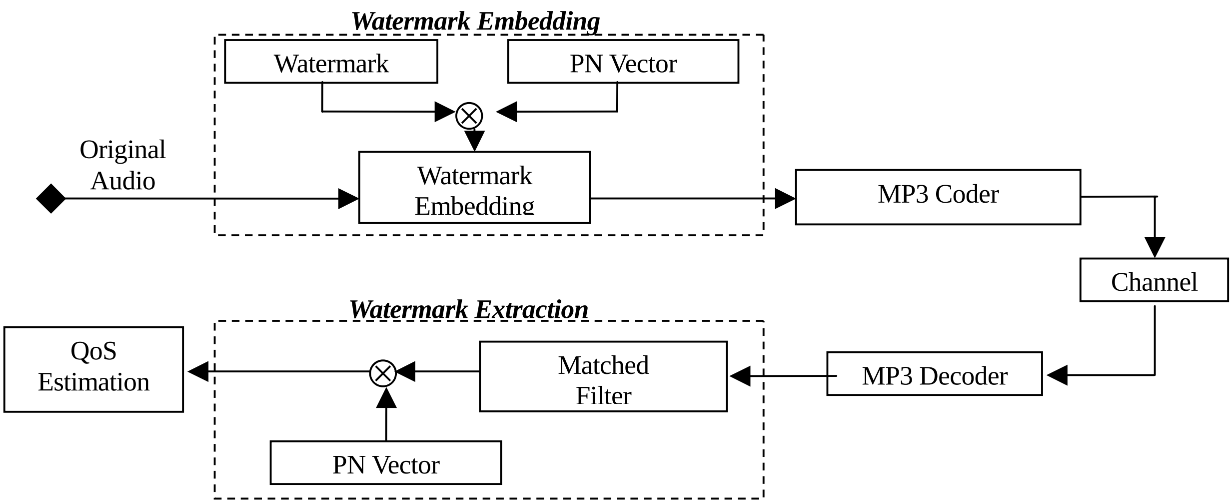

Fig. 1 - Block scheme of the audio watermarking signal processing for quality assessment of the received audio signal after an MP3 coding/transmission process

The result is that the temporal localization of the mark is different from an audio block to the other, so that the mark is an inaudible modification of the audio signal and, on the other hand, it is robust against permanent bit errors, due both to the physical network or its management (e.g. multi-path of the transmission channel, multi-user interference, excess loading factors).

3. OBJECTIVE QUALITY EVALUATION

In this Section, we show how it is effectively possible to provide a quality assessment of the received audio signal after the MP3 coding/transmission process, without affecting the quality of the communication. As stated before, the receiver implements audio decoding as well as watermark detection. At the receiving side, the watermark is extracted and compared with its original counterpart. In fact, after the MP3 decoding of the audio stream, a matched filter extracts the (known) watermark from the DCT of each *i*-th received watermarked block as follows:

$$\widehat{w}_i^{(s)}[k] = \widehat{G}_i^{(s)}[k] \cdot w[k] \qquad (3)$$

where $G^{\wedge}_i[k]$ is the DCT of the *i*-th received watermarked audio block and $w_i^{\wedge(s)}[k]$ is the spread version of the estimated watermark. The estimated watermark $w^{\wedge}_i[k]$ is matched to the reference one (despread with the *i*-th known PN vector) according to the following rule:

$$\widehat{w}_i[k] = \widehat{w}_i^{(s)}[k] \cdot p_i[k]\,. \qquad (4)$$

The matched filter is tuned to the particular embedding procedure, so that it can be matched to the randomly spread watermark only. The QoS is evaluated by comparing the extracted watermark with respect to the original one. In particular, the MSE is first evaluated for the *i*-th audio-block, and then it is averaged over the blocks under analysis, thus obtaining the following employed metric:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{K^2}\sum_{k=1}^{K}(w_i[k]-\widehat{w}_i[k])^2\right). \qquad (5)$$

It is worth noting that the metric (5), which is evaluated using the estimated watermarks over the *N* transmitted blocks, is employed to provide a quality assessment of the received audio signal after the MP3 coding/transmission process. In particular, such a QoS index can be usefully employed for a number of different purposes in mobile multimedia communications such as: control feedback to the sending user on the effective quality of the link; detailed information to the operator for billing purposes and diagnostic information to the operator about the communication link status.

Although sophisticated quality perceptual metrics could be used for quality assessment purposes in multimedia communications (see, for instance, [21]), the MSE between the estimated watermark and the original one is used in this contribution as a proof of concept of the objective method proposed here, and shown by extensive simulation trials in the next Section.

4. SIMULATION RESULTS

In this Section, some experimental results characterizing the effectiveness of the proposed method are presented. A wide number of simulation trials have been performed in order to validate the aforementioned hypotheses regarding the relationship between the degradations of the embedded watermark and the host signal. At the transmission side the QoS system described in the previous Sections of this work is embedded into the host audio signal. The watermarked audio is then compressed using an MP3 coder (MPEG-1 layer III coder) in order to obtain a coded bit-stream. The bit-stream is then transmitted over a noisy channel simulated by a Poisson's generator of random transmission errors, introducing a range of BER from 10^{-5} up to $5 \bullet 10^{-3}$. For sake of simplicity and without loss of generality, only audio tracks sampled at 44.1 KHz have been considered and we have tested the proposed QoS assessment setup for different compression ratios (64, 128, 192 kbps) of the MP3 coder. The watermark that has been employed is a low payload file (i.e. a binary sequence) modulated by the scaling coefficient α (i.e. the power of the mark). Increasing this value, the mark becomes more audible and a degradation of the audio signal occurs. On the contrary, by diminishing its value, the mark can be easily removed by the coder and/or channel's errors. Therefore, in the application scenario of our simulation trials, this coefficient has been chosen in such a way to compromise between the two aforementioned requirements ($\alpha = 0.04$, in all the following simulations). From the operating viewpoint, we are interested in the audio transmission over a given communication link determined by the (booked) maximum bit rate. In other words, the target quality is fixed by the negotiated channel capacity, while the actual quality depends also on the symbol errors introduced in the received data stream by the physical link, due to background noise as well as multi-path channel and interference effects. For this purpose, a number of simulations have been carried out. In particular, Figs. 2(a) and 2(b) report here the MSE of the estimated watermark with respect to the original one versus the bit error rate (BER) of the channel for MP3 coded audio signals (different compression ratios have been considered). It is worth noting that the MSE of the extracted watermark increases when the BER increases and the bit rate decreases. This is in accordance with the audible degradation that the audio suffers at increasing BER and decreasing bit rate. Moreover, as shown in Figs. 3(a) and 3(b), the quality degradation of the watermark embedded into the host signal has the same behavior of the one affecting the audio. It has to be noted that each curve of the graphs is normalized with respect to the maximum value, in order to compare both the two MSE.

The obtained results evidence the sensitivity of the watermarking quality index to the actual quality for given target quality levels and show the capability of this unconventional use of watermarking technique to trace the alterations suffered by the data through the communication channel, without affecting the quality of the communication. The proposed watermarking technique has been designed for application to mobile multimedia communication systems. In fact, such a QoS index can be usefully employed for a number of different purposes in mobile multimedia communications such as: control feedback to the sending user on the effective quality of the link; detailed information to the operator for billing purposes. It is worth pointing out that the mobile station (MS), that is the end-user of the communication process, needs to implement the evaluation of the provided QoS. As a consequence, the MS must perform real-time processing. Although actual MS does not perform hardware and/or software processing, the authors believe that in a few years, MS will host large processing capabilities because of the monotonically decreasing cost of very large scale integration. Therefore, the complexity of the QoS evaluation procedure, which is presented in this work, appears negligible in comparison with MP3 decoding.

5. CONCLUSIONS

In this work, we have proposed an *unconventional* use of an audio watermarking technique to provide a quality assessment of the received audio signal after an MP3 coding/transmission process, without affecting the quality of the communication. Although sophisticated quality perceptual metrics could be used for quality assessment purpose, the main scope of this contribution is to provide a proof of concept of the *objective* method, investigating its feasibility in the presence of audio music signals. Specifically, a fragile watermark is hidden in an MP3-like host data audio transport stream using a spread-spectrum approach. Since the alterations endured by the watermark are *likely* to be suffered by the MP3-file, as they follow the same communication link, the watermark's degradation can be used to estimate the overall alterations endured by the entire MP3-data. At the receiving side, the watermark is extracted and compared to its original counterpart. The QoS assessment is based on the evaluation of the mean-square-error between the estimated and the actual watermarks. The results obtained through our simulation trials confirm the validity of such approach. Thus, by knowing the end-to-end QoS, a number of different purposes can be reached in mobile multimedia communications such as: control feedback to the sending user on the effective quality of the link; detailed information to the operator for billing purposes and diagnostic information to the operator about the communication link status.

REFERENCES

[1] G. C. Langelaar, I. Setyawan, R.L. Lagendijk, "Watermarking digital image and video data. A state-of-the-art overview", *IEEE Sign. Proc. Mag.*, vol. 17, no. 5, pp. 20-46, Sept. 2000.

[2] C. I. Podilchuk, E. J. Delp, "Digital Watermarking: Algorithms and Applications", *IEEE Sign. Proc. Mag*., pp.33-46, July 2001.

[3] P. W. Chan, M. R. Lyu, R. T. Chin, "A Novel Scheme for Hybrid Digital Video Watermarking: Approach, Evaluation and Experimentation", *IEEE Trans. on Circuits and Systems for Video Techn.*, vol. 15, no. 12, pp.1638-1649, Dec. 2005.

[4] C.-S. Ko, K.-Y. Kim, R.-W. Hwang. Y.S. Kim, S.-B. Rhee, "Robust audio watermarking in wavelet domain using pseudorandom sequences", *4th ACIS Int. Conf. on Computer and Information Science*, pp.397 – 401, July 2005.

[5] J. Haitsma, T. Kalker, "Speed-change resistant audio fingerprinting using auto-correlation", *Proc. Int. Conf. on Acoustics, Speech, and Sign. Proc*., vol. 4, pp.728-731, Apr. 2003.

[6] A.N. Lemma, J. Aprea, W. OOmen, L. van de Kerkhof, "A Temporal Domain Audio Watermarking Technique", *IEEE Trans. on Sign. Proc*., vol. 51, no.4, pp. 1088-1097, Apr. 2003.

[7] P. Campisi, M. Carli, G. Giunta and A. Neri, "Blind Quality Assessment System for Multimedia Communications Using Tracing Watermarking", *IEEE Trans. on Sign. Proc*., vol. 51, no. 4, pp. 996-1002, Apr. 2003.

[8] D. Kirovski, and H. Malvar, "Robust Spread-Spectrum Audio Watermarking," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, May 2001, pp. 1345–1348.

[9] M. Arnold, "Audio Watermarking: Features, Applications and Algorithms," *Proc. of the IEEE Int. Conf. on Multimedia and Expo (ICME 2000)*, New York, July 2000, pp. 1013–1016.

[10] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.

[11] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A DCT-domain system for robust image watermarking," *Sign. Process*., vol. 66, no. 3, pp. 357–372, May 1998.

[12] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images", *IEEE Trans. Image Proc.*, vol 6, no 12, pp. 1621-1633, Dec. 1997.

[13] S. Wu, J. Huang, D. Huang; Y. Q. Shi, "Efficiently self-synchronized audio watermarking for assured audio data transmission", *IEEE Trans. on Broadcasting*, vol. 51, no. 1, pp. 69-76, March 2005.

[14] C.-T. Wang, T.-S. Chen, W.-H. Chao, "A new audio watermarking based on modified discrete cosine transform of MPEG/audio layer III", *IEEE Int. Conf. on Networking, Sensing and Control*, vol. 2, pp. 984 – 989, 2004.

[15] S. Esmaili, S. Krishnan, K. Raahemifar, "A novel spread spectrum audio watermarking scheme based on time-frequency characteristics", *IEEE. Canadian Conf. on Electrical and Comp. Eng., 2003*, vol. 3, pp. 1963–1966, May 2003.

[16] I.-K. Yeo, H. J. Kim, "Modified patchwork algorithm: a novel audio watermarking scheme", *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 4, pp. 381-386, July 2003.

[17] P. Arttameeyanant, P. Kumhom, K. Chamnongthai, "Audio watermarking for Internet", *IEEE Int. Conf. on Industrial Technology*, vol. 2, pp. 976-979, Dec. 2002.

[18] M. Arnold, "Audio watermarking: features, applications and algorithms", *IEEE Int. Conf. on Multimedia and Expo*, vol. 2, pp. 1013-1016, 2000.

[19] D. Pan, "Tutorial on MPEG/Audio Compression" *IEEE Multimedia Journal*, Summer 1995 issue, May 1995.

[20] ISO11172-3 ISO/IEC: International Standard Information Technology, Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits – Part 3: Audio, ISO11172-3.1996.

[21] S. Winkler, "Visual fidelity and perceived quality: Toward comprehensive metrics," *Proc. SPIE*, vol. 4299, 2001.

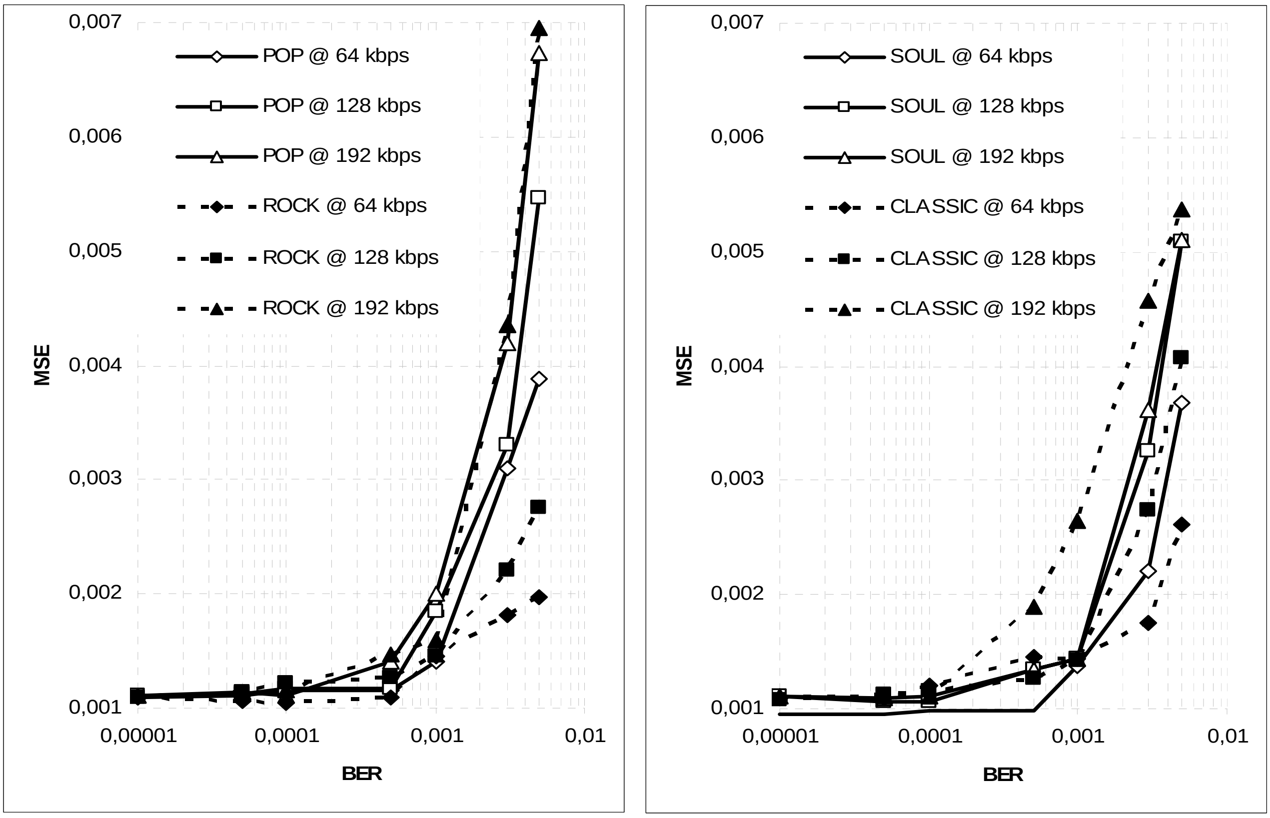

Fig. 2. MSE of the watermark extracted from MP3 audio signals: (*left*) POP and ROCK music, (*right*) SOUL and CLASSIC music versus the BER of the channel for different compression ratios of the MP3 (MPEG-1 layer III) coder.

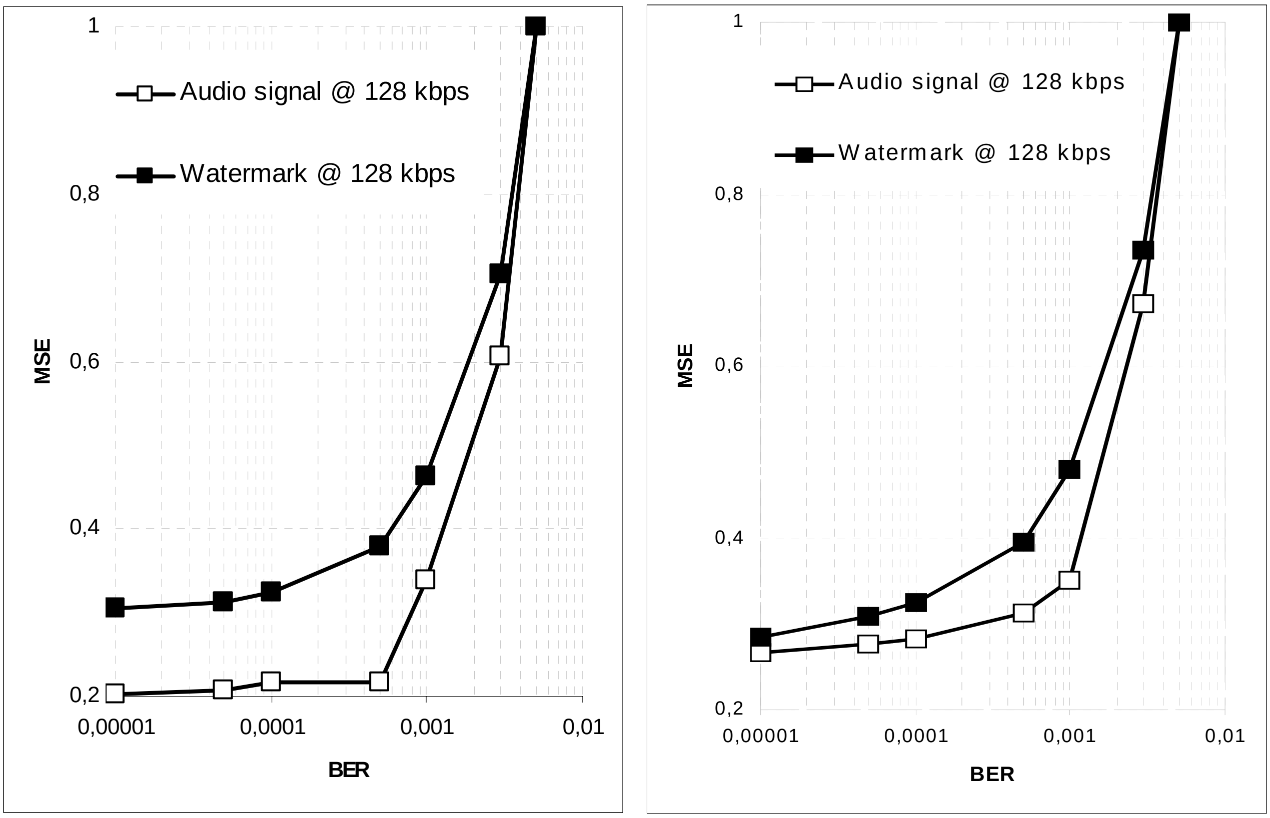

Fig. 3. MSE (normalized to 1) of the watermarked MP3 signal and of the extracted watermark: (*left*) POP music; (*right*) CLASSIC music, evaluated at a compression ratio of 192 kbps.