

Design of Audio Digital Watermarking System Resistant to Removal Attack

Valery Korzhik, IEEE Member, and Vasily Alekseev
The Bonch-Bruевич Saint-Petersburg
State University of Telecommunication
Saint-Petersburg, Russia
Email: val-korzhik@yandex.ru, i@vasay.ru

Guillermo Morales-Luna
Computer Science
CINVESTAV-IPN
Mexico City, Mexico
Email: gmorales@cs.cinvestav.mx

Abstract—We consider a digital watermarking system intended for an embedding of additional information into audio (typically musical) files that should be resistant against a removal attack. The proposed embedding procedure is based on a reverberation and extraction procedure executing a cepstral transform. A removal attack based on blind dereverberation is investigated both theoretically and experimentally. In order to prevent such an attack, a slight modification of the embedding procedure is also proposed. Further experiments show that the proposed watermarking system provides both a good quality of the cover audio-signal and a sufficiently large embedding rate.

Index Terms—Audio watermarking, blind dereverberation, cepstrum, reverberation

I. INTRODUCTION

IT is well known that the technology of digital watermarking (WM) is the most effective approach to provide copyright protection for digital media products. Examples of such products are the digital audio and video works. In the current paper we consider audio works (first of all musical files presented as digital signals in format *wav*). Such objects in which it is necessary to embed an additional information will be called in the sequel *cover objects* (CO). Dishonest users (pirates indeed) may try to remove the embedded WM without remarkable corruption of the CO hoping to illegally redistribute them to other users. They may accomplish the desired result after some processing of the watermarked objects in such a way that the legal users were unable to extract WM correctly from the redistributed copies and consequently they will be unable to perform a forensic consideration against pirates.

On the contrary, the owners of the products may try to embed into the CO a WM that cannot be removed without significant corruption at the CO. A significant corruption of CO results in their lower values at the market and a redistribution occurs useless.

Several embedding WM techniques for audio-signals are well known and they have been extensively used, e. g. *phase-shift-keying* (PSK) modulation [1] or WM system based on *echo hiding* (EH) [2]. But as it was shown [2], [3] that within both PSK and EH WM systems the embedded WM can be easily removed without significant degradation of the CO.

The use of spread spectrum signals in the embedding procedures that are controlled by a secret *stegokey* seems to

be very attractive. But a more careful consideration [4] shows that such signals are vulnerable to desynchronization attacks.

At a single glance, the use of a reverberation procedure with a secure pulse response of the reverberation filter, controlled by a stegokey, is the best approach. In fact, on the one hand the use of a reverberation with filter pulse response close to a *room pulse response filter* provides a good quality of audio CO [5]. On the other hand, the use of complex pulse response forms prevents a compensation of reverberation (making a dereverberation – in other words) that could be allow to remove the embedding.

But unfortunately, a changing of pulse response form on every bit interval results (as our experiments showed) in a significant corruption of CO. Therefore we propose some “intermediate” approach that is presented in Section II. But without some additional transforms, described in Section IV, the WM system presented in Sections II and III will be yet vulnerable to the blind dereverberation attack described in those sections also. The proposed modified WM system is presented in Section IV. Section V concludes the paper and presents some open problems for the future work.

II. ATTACK ON A WM SYSTEM THAT IS BASED ON THE EMBEDDING WITH A REVERBERATION USAGE

Let us assume that a given WM system uses some fixed (but sufficiently complex) reverberation *filter pulse response* $(h_b(n))_{n=1}^N$ for all watermarking session, where N is the number of samples on every bit interval. In order to embed bits $b = 0$ or $b = 1$ it is used only fixed but different time delays with each filter corresponding to additional information. Then the digital WM-ed signal $(Z(n))_{n=1}^N$ on each bit interval can be presented as follows:

$$\forall n = 1, \dots, N : Z(n) = S(n) * h_b(n) \quad , \quad b \in \{0, 1\} \quad (1)$$

where $(S(n))_{n=1}^N$ is the input audio signal (CO), $(h_b(n))_{n=1}^N$ is the filter pulse response depending on the embedding bit b , $*$ is the operation of convolution, and N is the number of samples on each symbol interval. By applying the cepstrum transform to both sides of (1) we get [6]:

$$\forall n = 1, \dots, N : \tilde{Z}(n) = \tilde{S}(n) + \tilde{h}_b(n) \quad , \quad b \in \{0, 1\} \quad (2)$$

where $\tilde{\cdot}$ denotes the *cepstrum transform*:

$$C(x)(n) = \frac{1}{N} \sum_{k=0}^{N-1} e^{\frac{2\pi}{N} \iota nk} (\iota \Theta(k) + \log x'(k)) = \tilde{x}(n) \quad (3)$$

with

$$\forall k = 1, \dots, N : x'(k) = \sum_{m=0}^{N-1} e^{-\frac{2\pi}{N} \iota mk} x(m),$$

$(x'(k))_{k=1}^N$ is the signal amplitude, $(\Theta(k))_{k=1}^N$ is the signal phase and $\iota = \sqrt{-1}$.

In reality, relation (2) is only an approximation of a finite signal. The accuracy of expression (2) depends on the number of zeros added to the finite signal. If the number of added zeros is sufficiently large, then relation (2) holds with small errors. The advantage of (2) compared with expression (1) consists in the easiness of the cepstrum transform to apply well known algorithms for optimal receivers [7] if the interference $(S'(n))_{n=1}^N$ can be approximated by white Gaussian noise.

The extraction algorithm for such a WM system is the well known *correlation receiver*:

$$b = \text{Arg max}_{b \in \{0,1\}} \sum_{n=1}^N \tilde{Z}(n) \tilde{h}_b(n). \quad (4)$$

Let us assume that an attacker that trying to remove the WM is able to estimate somehow the filter cepstrum pulse responses for each $b \in \{0,1\}$ as $(\tilde{h}'_b(n))_{n=1}^N$ on each of bit interval. Then an attack intended to remove WM could be:

$$\forall n = 1, \dots, N : \tilde{Z}_b(n) = C^{-1} (\tilde{Z}(n) - \tilde{h}'_b(n)) \quad (5)$$

where C^{-1} is the inverse of the cepstrum transform C given in (3). (In favor of the attacker, we do not consider here the hardness to perform the transform C^{-1} .) It is worth to note that an operation to remove a reverberation from an audio signal is called *blind dereverberation*. This problem was investigated in many papers [8], [9], [10], [11], [12], [13] and others. But the goal of such signal transform was to make the audio signal free from additional reverberation interference that may occur in a natural manner.

In our case, it is not sufficient to make the audio signal sufficiently free from reverberation just “by ear”. We require to make impossible WM extraction from the dereverberated signal even with the use of an optimal receiver. Moreover, for the purpose of dereverberation removal there were used multiple microphones placed on some distances one against another [10]. Of course such approach cannot be used in our scenario.

Let us estimate the error probability P (incorrect bit b extraction) for the WM system owner using the decision rule (4) where

$$\forall n = 1, \dots, N : \tilde{Z}_a(n) = \tilde{Z}(n) - \tilde{h}'_b(n).$$

It is easy to see from (2), (4) and (5) that even for opposite signals $\tilde{h}'_0(n)$ and $\tilde{h}'_1(n)$,

$$\begin{aligned} P &= \Pr(1|0) \\ &= \Pr \left(\xi \leq - \sum_{n=1}^N (\tilde{h}_0(n) - \tilde{h}'_0(n)) \tilde{h}_0(n) \right) \end{aligned} \quad (6)$$

with

$$\xi = \sum_{n=1}^N \tilde{S}(n) \tilde{h}_0(n).$$

After a changing of variables we get from (6),

$$P = \frac{1}{\sqrt{2\pi\sigma^2\tilde{A}}} \int_{-\infty}^{\tilde{A}} \exp \left(-\frac{x^2}{2\sigma^2\tilde{A}} \right) dx \quad (7)$$

where $\tilde{A} = \sum_{n=1}^N (\tilde{h}_0(n) - \tilde{h}'_0(n)) \tilde{h}_0(n)$, $A = \sum_{n=1}^N \tilde{h}_0^2(n)$ and $\sigma^2 = \text{Var}(\tilde{S}(n))$. (We note that relation (7) holds whenever σ is a zero mean Gaussian sequence with variance $\sigma^2 A$.) It is easy to prove that

$$\tilde{A} = A(1 - \eta) \quad (8)$$

where $\eta = \frac{1}{\tilde{A}} \sum_{n=1}^N \tilde{h}'_0(n) \tilde{h}_0(n)$. Substituting (8) into (7) we get after a simple transform

$$P = 1 - F \left(\sqrt{\frac{A(1 - \eta)^2}{\sigma^2}} \right) \quad (9)$$

where

$$\forall x \in \mathbb{R} : F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left(-\frac{t^2}{2} \right) dt.$$

(If the signals $\tilde{h}'_0(n)$ and $\tilde{h}'_1(n)$ are not opposite, then equality (9) holds as a lower bound (in favour of the attacker).

We see from (9) that if $\eta = 0$, namely there is a bad estimation of $\tilde{h}'_0(n)$, then the attack occurs in an inefficient way. But if $\eta = 1$, there results in $P = \frac{1}{2}$, which means a “break of the legal WM channel”. Then the estimation attack is effective because it removes completely the WM embedding.

In Fig. 1 there are shown the dependencies of the legal user error symbol probability calculated by (9) against of parameter η for different values of $\frac{A}{\sigma^2}$.

We see from the dependencies presented in Fig. 1 that in order to provide high efficiency in the attack it is necessary to get the parameter η close to the value 0.8. Hence, an attacker should correctly estimate the filter pulse responses of legal user. We note first of all that such problem cannot be solved exhaustively over all possible filter pulse response wave forms.

In fact, the typical length of “room pulse” that keeps a good quality of a musical file after embedding, is about 180 samples. Assuming that the pulse response amplitude is at most around 0.2 with respect to audio signal amplitude, we get for a total number of quantization levels 65536 for the format wav, and the number of acceptable levels for pulse response will be about 13107. Then a set of all possible pulse response wave forms appears with cardinality around 1.4×10^{741} , which is certainly an untractable value for an exhaustion attack.

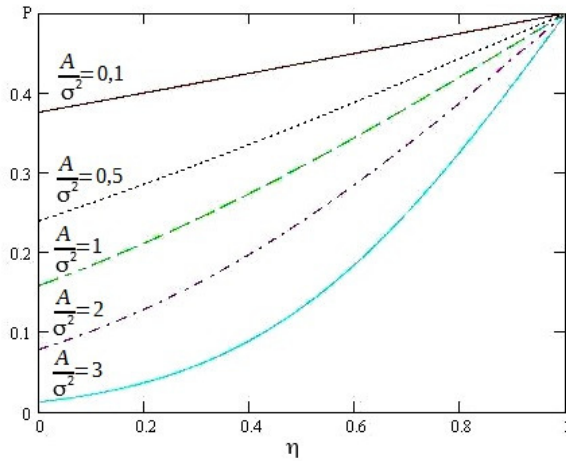


Fig. 1. The dependencies of the legal user error symbol probability P against η for different $\frac{A}{\sigma^2}$.

If we assume that the filter pulse responses $\tilde{h}'_0(n)$ and $\tilde{h}'_1(n)$ differ only by a fixed and known delay N_0 then the attacker could find all bit intervals I_0 corresponding to $b = 0$ and I_1 corresponding to $b = 1$.

Next it is possible to average separately all cepstrum corresponding wave forms in order to get an approximation of the cepstrum pulse response as follows:

$$\begin{aligned} & \frac{1}{L} \left[\sum_{n \in I_0} \tilde{Z}(n) + \sum_{n \in I_1} T_{N_0}(\tilde{Z}(n)) \right] \\ &= \tilde{h}_0(n) + \frac{1}{L} \sum_{i=1}^L \tilde{S}_i(n) \end{aligned} \quad (10)$$

Using (10) and the expression of η in (8) we get

$$\eta = 1 - \frac{\sum_{n=1}^N \tilde{h}_0(n) \frac{1}{L} \sum_{i=1}^L \tilde{S}_i(n)}{\sum_{n=1}^N \tilde{h}_0^2(n)} = 1 - \varepsilon. \quad (11)$$

Let us find the variance of the random variable ε assuming that $\text{Var}(\tilde{S}_i(n)) = \sigma^2$ and the samples of cepstrum $\tilde{S}_i(n)$ are *i.i.d.* random values. We can write

$$\begin{aligned} \text{Var}(\varepsilon) &= \frac{\text{Var} \left(\sum_{n=1}^N \tilde{h}_0(n) \frac{1}{L} \sum_{i=1}^L \tilde{S}_i(n) \right)}{\sum_{n=1}^N \tilde{h}_0^2(n)} \\ &= \frac{\sigma^2}{L \sum_{n=1}^N \tilde{h}_0^2(n)} = \frac{\sigma^2}{LA} \end{aligned} \quad (12)$$

Next we can use relation (12) for known cepstrum pulse response $\tilde{h}_0(n)$ and known parameters L and σ^2 in order to estimate that the parameter η is at most $3 \text{Var}(\varepsilon)$ with probability 0.997.

a) *Example:* Assume $\frac{A}{\sigma^2} = \frac{1}{2}$, $L = 360$, then, by (12), $\text{Var}(\varepsilon) < 6 \times 10^{-4}$ and the parameter η is at least 0.98 with the probability 0.997. Then, from Fig. 1, we see that for the attack estimation presented above the extracted bit error probability for legal user occurs close to 0.5 and hence this attack be very effective. \square

But a gap in the attack estimation is the fact that so far it is unknown how an attacker could be able to find all bit intervals belonging separately to the embedding of bits.

Since the forms of filter pulse responses are constant for different bit intervals (in line with our previous assumption) and they differ only within a fixed delay, the same situation appears for the corresponding cepstrums. Thus, if for a pair of bit intervals I_i and I_j corresponding to equal bits b and \tilde{b} , $b = \tilde{b}$, then the following crosscorrelation for the corresponding cepstrum wave forms $\tilde{Z}_i(n)$ and $\tilde{Z}_j(n)$ is obtained:

$$\begin{aligned} \Lambda &= \frac{1}{N} \sum_{n=1}^N \tilde{Z}_i(n) \tilde{Z}_j(n) \\ &= \frac{1}{N} \sum_{n=1}^N \left(\tilde{S}_i(n) + \tilde{h}_b(n) \right) \left(\tilde{S}_j(n) + \tilde{h}_b(n) \right) \\ &= \frac{1}{N} \sum_{n=1}^N \left(\tilde{S}_i(n) \tilde{S}_j(n) + \tilde{S}_i(n) \tilde{h}_b(n) + \right. \\ & \quad \left. \tilde{h}_b(n) \tilde{S}_j(n) + \tilde{h}_b(n) \tilde{h}_b(n) \right). \end{aligned} \quad (13)$$

For the case of different embedding on the i -th and j -th bit intervals, that is, $b \neq \tilde{b}$, we get

$$\begin{aligned} \Lambda' &= \frac{1}{N} \sum_{n=1}^N \tilde{Z}_i(n) \tilde{Z}_j(n) \\ &= \frac{1}{N} \sum_{n=1}^N \left(\tilde{S}_i(n) + \tilde{h}_b(n) \right) \left(\tilde{S}_j(n) + \tilde{h}_{\tilde{b}}(n) \right) \\ &= \frac{1}{N} \sum_{n=1}^N \left(\tilde{S}_i(n) \tilde{S}_j(n) + \tilde{S}_i(n) \tilde{h}_{\tilde{b}}(n) + \right. \\ & \quad \left. \tilde{h}_b(n) \tilde{S}_j(n) + \tilde{h}_b(n) \tilde{h}_{\tilde{b}}(n) \right) \end{aligned} \quad (14)$$

By comparing equations (13) and (14) we conclude that in the first case Λ is larger than Λ' in the second case. Therefore we may select a threshold and to decide that the i -th and the j -th interval correspond to the same bit interval, $b = \tilde{b}$, if the threshold is exceeded and, otherwise, they correspond to different bit intervals, $b \neq \tilde{b}$. Thus it is possible to find the sets I_0 and I_1 for the calculation in (10).

However another question arises: how can an attacker find filter pulse response but not filter cepstrum pulse response using (10)?

It has been proved in [11] that for small embedding amplitude it is possible to take into account only the first term in

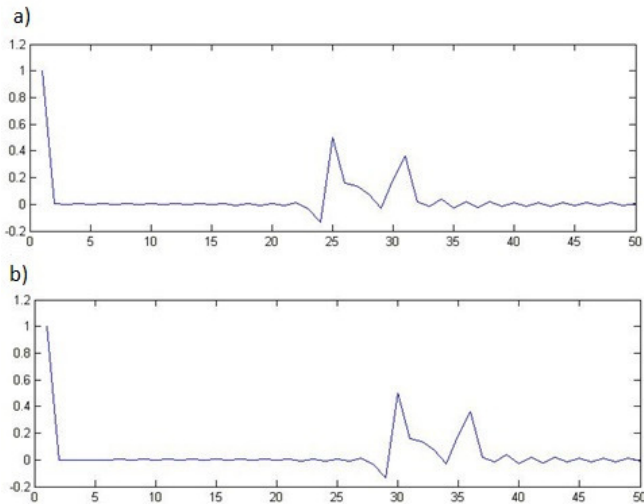


Fig. 2. Filter pulse response: a) for bit "1", b) for bit "0".

the Taylor series for the cepstrum expansion of signal in (2). This means that the last equation can be rewritten as

$$\tilde{Z}(n) = \tilde{S}(n) + \lambda h'_b(n) \quad (15)$$

where λ is some scale coefficient, $h'_b(n)$ is the already filter pulse response but not the cepstrum pulse response in (2).

Expression (15) asserts that if an attacker has estimated correctly the cepstrum pulse response, then he (or she) will be able to find the pulse response after a specification (maybe even though an exhaustive trial) of the coefficient λ .

After the full calculation of the filter pulse responses, an attacker, with the knowledge of bits embedding on each bit interval, may manage to apply the inverse filter pulse response and consequently to remove all the embedded information.

However, in the above theoretical investigation a model for the cover objects unavailable in practice has been suggested. Therefore in the next section we investigate experimentally the proposed attack. In Section IV we modify the embedding scheme in such a way that it will be resistant against the proposed attack.

III. EXPERIMENTAL INVESTIGATION OF THE PROPOSED DEREVERBERATION ATTACK

We select the filter pulse response (FPR) for both embedding bits $b = 0$ and $b = 1$ shown in Fig. 2. The chosen delays for embedding are 30 and 25 samples for bits zero and one, respectively. Cepstrum of these FPR are shown in Fig. 3. These figures confirm the assertions given before that firstly cepstrum delays coincide with FPR delays and secondly, that cepstrum wave forms copy FPR wave forms.

All bit intervals corresponding to bits $b = 0$ and $b = 1$ were found with the use of the crosscorrelation Λ , and Λ' given by eq's. (13), (14) respectively.

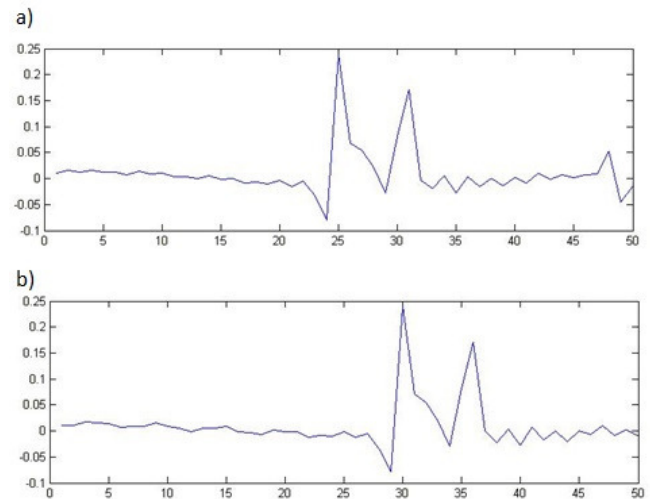


Fig. 3. Filter cepstrum pulse response: a) for bit "1", b) for bit "0".

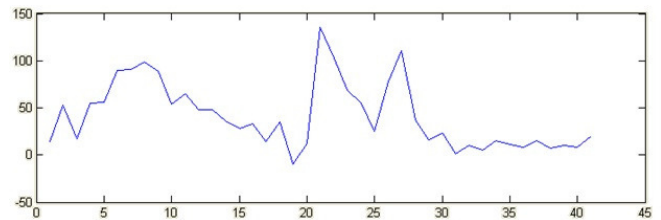


Fig. 4. Averaged FCPR in line with eq. (10).

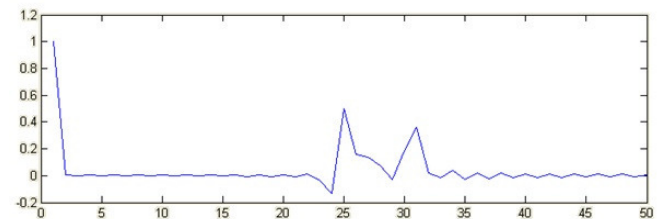


Fig. 5. Estimation of the FPR after a selection of a scale factor.

In Fig. 4 the averaged FCPR is presented in line with eq. (10). We see from this figure that a form of FCPR copies a form of FPR up to some scale factor.

If an attacker is able to find the scale factor then the wave form of the FPR can be easily estimated (see Fig. 5).

Now, the dereverberation attack can be performed with the following steps:

- 1) For a known FPR (Fig. 5) calculate the FCPR (see Fig. 6)
- 2) Reflect with respect to zero the wave form of FCPR
- 3) Find FPR for the attack filter computing inverse cep-

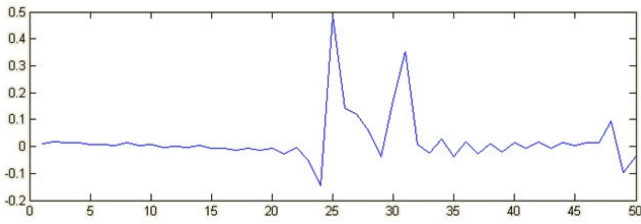


Fig. 6. The FCPR calculated from the FPR given in Fig. 5.

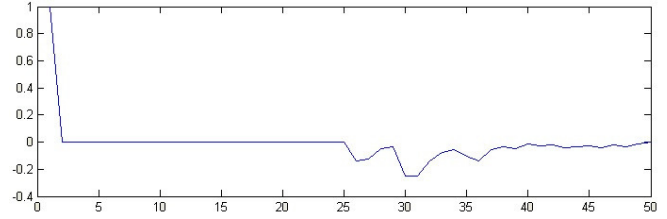


Fig. 8. FPR wave form obtained by (16).

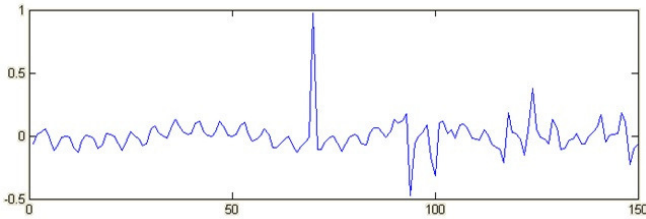


Fig. 7. The FPR for dereverberation attack.

strum transform from FPR. The result is presented in Fig. 7. In a similar manner, there can be calculated the inverse FPR for the embedding of the bit $b = 0$.

- 4) Apply the inverse filters to the embedded bits “0” and “1” which have been found before in the corresponding bit intervals.
- 5) Use the transition function between bit intervals with linear form that is necessary to keep high quality of audio signal after dereverberation procedure.

In Table I the extracted bit error probabilities before and after dereverberation attack under different parameters of WM system are presented. The wave forms of FPR were presented in Fig. 9. They have finite length equal to 180 samples. We see from this table that before attack the proposed WM system is working acceptably but after the dereverberation attack the bit error probability is close to 50%, that is similar to “break of channel”. (We note the fact that sometimes the probability exceeded 50% owing to an incorrect estimation of scale factor. But it does not affect on our conclusion.)

IV. MODIFICATION OF THE WM SYSTEM TO BE RESISTANT AGAINST A DEREVERBERATION ATTACK

In order to protect the WM system from the above proposed dereverberation attack it is necessary to make impossible for an attacker to separate 0-bit intervals from 1-bit intervals.

In fact, if an attacker does not know which bit intervals correspond to the embedding bit “1” and which ones to the bit “0”, then by replacing expression (10) to a summation over all the bit intervals,

$$\Lambda'' = \frac{1}{N} \sum_{n=1}^N \tilde{Z}(n) \tag{16}$$

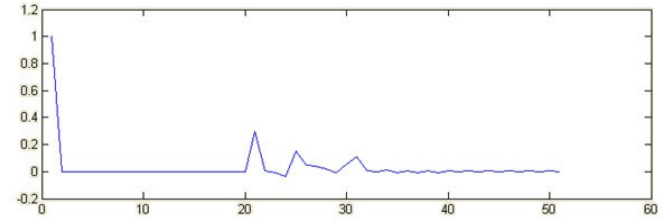


Fig. 9. Wave form of FPR with additional pulse on 21-th sample of bit interval.

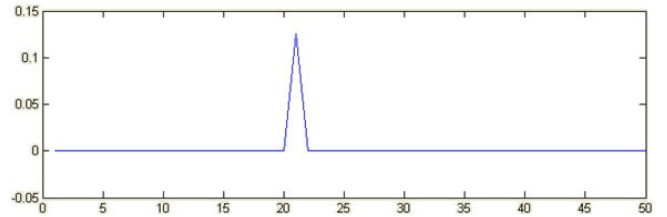


Fig. 10. Result of crosscorrelation computation with additional pulse on the 21-th sample.

a large corruption of FPR wave form in comparison with original one results. In Fig. 8 a FPR wave form is presented after such “total averaging”

We see that the FPR in Fig. 8 has no any similarity with the original FPR wave form (see Fig. 2), hence an attacker will be unable to arrange a dereverberation attack. (In fact we have checked that the use of such FPR in a dereverberation attack cannot even result in a remarkable increasing of the extracted bit errors.)

In order to prevent a crosscorrelation attack (13), we propose to add to the WM signal short pulses at the beginning of each bit interval. (See Fig. 9 where an additional pulse is presented on the 21-th samples of the bit interval).

The use of the crosscorrelation attack given by (13), (14), results in an occurrence of a single pulse independently on whether there is a coinciding or a discrepancy among the information bits corresponding to signal $\tilde{Z}_i(n)$ and $\tilde{Z}_j(n)$ (see Fig. 10) for a confirmation).

Thus we can conclude that a modification of the reverberation-based WM system by additional pulses results

TABLE I
THE EXTRACTED BIT ERROR PROBABILITIES BEFORE AND AFTER DEREVERBERATION ATTACK FOR DIFFERENT SYSTEM PARAMETERS.

Name of music files and their duration	Delays of WM signal		The length of bit intervals (in number of samples)	The number of the embedded bits	Bit error rate before attack in %	Bit error rate after attack in %
	1	0				
Vysocki "Song of Boxer" (fragment 20 sec)	25	29	4000	142	4.5%	72%
Vysocki "Song of Boxer" (fragment 20 sec)	25	29	6000	94	0%	78%
Vysocki "Song of Boxer" (fragment 20 sec)	15	19	6000	94	17%	63%
Yuta, "Jealousy" (fragment 29 sec)	25	29	10000	55	2%	48%
Yuta, "Jealousy" (fragment 29 sec)	25	29	5000	113	1%	57%
Yuta, "Jealousy" (fragment 29 sec)	20	24	5000	113	7%	61%

in a resistance of this system to a most power blind dereverberation attack.

We have tested the proposed WM system also with respect to audio signal quality after embedding. A group consisting of 5 experts has come into a conclusion that a quality of musical files after WM embedding keeps practically the same as the embedding before.

V. CONCLUSION

In this paper an audio WM system resistant to a remove attack is proposed. The embedding of WM in this system is performed by a reverberation of audio signal that is controlled by a secret stegokey. The main advantage of the reverberation based watermarking system is its possibility to provide a high quality of audio signal after embedding. But there exists an effective attack for such WM system known as blind dereverberation attack. We investigated this attack in detail and showed that in fact it is able to remove the embedding information without significant degradation of audio signal quality. Therefore we propose some modification of WM-based system and show that then such attack is useless.

Experimental investigation confirm our conclusion. This system can be practically applied to copyright purposes. It would be interesting in the future to investigate more sophisticated attack on WM-based system although maybe with some degradation of audio signal quality.

REFERENCES

- [1] M. Arnold, P. G. Baum, and W. Voelbing, "Information hiding," S. Katzenbeisser and A.-R. Sadeghi, Eds. Berlin, Heidelberg: Springer-Verlag, 2009, ch. A Phase Modulation Audio Watermarking Technique, pp. 102–116, DOI: 10.1007/978-3-642-04431-1_8. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-04431-1_8
- [2] V. I. Korzhik, G. Morales-Luna, and I. Fedyanin, "Audio watermarking based on echo hiding with zero error probability," *International Journal of Computer Science and Applications*, vol. 10, no. 1, pp. 1–10, 2013.
- [3] V. Alekseyev, A. Grudinin, and V. Korzhik, "Design of robust audio watermark system," in *Proceedings of the XI International Symposium on Problems of Redundancy in Information and Control Systems*, Aug 2007, pp. 163–165.
- [4] H. Liu and W. Zhang, "Overview of audio watermarking algorithm against synchronization attacks," in *Advances in Intelligent Systems Research: ICAITA-16*, Aug 2016, DOI: 10.2991/icaita-16.2016.52.
- [5] J. M. Arend and C. Pörschmann, "Audio watermarking of binaural room impulse responses," in *Audio Engineering Society Conference: 2016 AES International Conference on Headphone Technology*, Aug 2016, DOI: 10.17743/aesconf.2016.978-1-942220-09-1. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=18346>
- [6] J. Proakis, *Digital Communications, Fourth Edition*. Mc Graw Hill, 2001.
- [7] D. G. Childers, D. P. Skinner, and R. C. Kemerait, "The cepstrum: A guide to processing," *Proceedings of the IEEE*, vol. 65, pp. 1428–1443, 1977, DOI: 10.1109/PROC.1977.10747.
- [8] T. Nakatani, M. Miyoshi, and K. Kinoshita, "One microphone blind dereverberation based on quasi-periodicity of speech signals," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds. Cambridge, MA: MIT Press, 2003, p. None. [Online]. Available: http://books.nips.cc/papers/files/nips16/NIPS2003_SP06.pdf
- [9] C. Evers, "Blind dereverberation of speech from moving and stationary speakers using sequential Monte Carlo methods," Ph.D. dissertation, The University of Edinburgh (United Kingdom), 2010.
- [10] H. Attias, J. Platt, A. Acero, and L. Deng, "Speech denoising and dereverberation using probabilistic models," November 2000. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/speech-denoising-and-dereverberation-using-probabilistic-models/>
- [11] N. Cvejic and T. Seppanen, *Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks*. Hershey, PA, USA: IGI Global, 2007, DOI: 10.4018/978-1-59904-513-9.
- [12] G. Chardon, T. Nowakowski, J. de Rosny, and L. Daudet, "A blind dereverberation method for narrowband source localization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 815–824, Aug 2015, DOI: 10.1109/JSTSP.2015.2422673.
- [13] K. Imoto and N. Ono, "Spatial cepstrum as a spatial feature using a distributed microphone array for acoustic scene analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1335–1343, 2017, DOI: 10.1109/TASLP.2017.269059.