TIC-TAC, FORGERY TIME HAS RUN-UP! LIVE ACOUSTIC WATERMARKING FOR INTEGRITY CHECK IN FORENSIC APPLICATIONS

V.A. Niță, A. Ciobanu

Politehnica University of Bucharest, Telecommunications Department, Bucharest, Romania

ABSTRACT

A common problem in audio forensics is the difficulty to authenticate an audio recording. In this paper we provide a novel and reliable solution to this problem by making use of a control signal, visible and audible on the actual recording, yet ignored by the listener, the TIC-TAC signal. We describe our live watermark solution, we incorporate it in an integrity check algorithm and we provide meaningful preliminary tests. Their results, computed in terms of precision show an outstanding performance: 100% detection rate for edited recordings by means of deleting/inserting audio fragments longer than 5ms and 0% false alarm rate for unedited recordings.

Index Terms— Audio Live Watermarking, Integrity Protection, Zero Delay

1. INTRODUCTION

THESE days, audio forgery can be performed by almost anyone with minimum audio editing knowledge due to the easy access to high-end audio editing software (e.g. Adobe Audition, Sound Forge, AVS Audio Editor). Having access to such software equipped with friendly and intuitive user interfaces, it becomes clear why the integrity of an audio recording is constantly challenged in a court of law. Audio forensics develops in two main directions: the audio authenticity check (i.e. authenticating the recording time [1], the context of the recording [2] and the acquisition sensor [3]), and the audio integrity check [4] (i.e. the ability to detect if an ill-intentioned person has tried to change the meaning of the message contained in an audio recording).

Audio integrity check implies searching for particular artifacts caused by the deletion and/or insertion of audio fragments in the recording or by the collage of fragments extracted from different audio sources (including the original source). The common artifacts prone to appear are: discontinuities in the waveform of the recorded signal [5], double compression artifacts [6], artifacts in the background noise [7], artifacts in the room fingerprint [8], artifacts induced by the nonlinearities of the recording microphone [3] and artifacts visible in residual (spectral) components, such as the hum noise [4] (widely known as the electric network frequency – ENF criterion).

We divide the methods used for integrity check into two categories: passive and active methods.

Passive integrity check methods assume that artifacts appear in the audio recording due to the editing process (e.g. double compression, discontinuities, etc.). The artifacts generated in this manner may be hard to find, especially if the person altering the audio recording is aware of them (for example adding noise over the edited fragments hides the discontinuities [9]). The accuracy of these methods can be too low, making it difficult to promote them as reliable audio forensic instruments and accept their outcome in a court of law.

Active methods enhance the integrity check by adding a residual signal in the recording process (which is equivalent to adding a watermark over the original signal). Hence, when a person alters the audio data, we are able to determine its integrity in a more reliable way by inspecting the changes in the artificially added signal (the residual).

A criterion used with reliable results [4] is the ENF criterion, which as an integrity check method can be placed somewhere between the active and the passive methods. This is due to the fact that while it relies on a residual signal (incorporated in the recording prior to forgery), this signal appears as a consequence of connecting the power source of the recorder to the power grid and not as an intended watermark. In this case the resulted residual signal is the hum noise [1], which is an artificial, *unwanted* signal affecting the audio quality of the recording, but very helpful in forensics applications. For instance, the insertion or the deletion of audio fragments in a recording can be detected by searching for discontinuity jumps in the phase spectrum of this signal. Further details on this topic can be found in [4]. However the problem with the ENF criterion is that this residual signal may not be present at all times, especially if the recorder is battery powered. Another important issue concerns the easiness with which the hum noise can be removed since its spectrum occupies a maximum bandwidth of only 2Hz, which lies in the lower frequency band, where the speech signal has no significant components. Therefore, if one should remove the hum noise, it would in fact enhance the audio signal, and not distort it.

In this paper we propose a new live watermarking algorithm inspired by the ENF criterion [1] and by a Sonic live watermarking method used for live performances [10].

The paper is further organized as follows. Section 2 provides an introduction to the concept of audio watermarking for live performances, while Section 3 offers the details of the proposed method. Section 4 brings forth an algorithm used to check the integrity of an audio recording, based on the live TIC-TAC watermarking. Objective tests validate our algorithm and reveal top performance. Finally Section 5 concludes the paper.

2. LIVE AUDIO WATERMARKING FOR FORENSICS

Typical audio watermarking involves hiding a message (referred to as watermark) in the audio data, which will later serve as a mean of copyright protection and authentication. The watermark is generally used for stored digital data (offline applications) and not for broadcast applications. However in [10]

the authors proposed an idea of a real-time watermark in order to extend the concept to broadcasts and real-time applications.

The performances criteria of an offline watermark algorithm are [11]:

-**Imperceptibility**, the watermark should be masked by the audio data so that a listener is not able to hear it;

-**Robustness**, the ability to extract the watermark even if the system has been affected by malicious integrity attacks;

-**Security**, only authorized persons should be able to access the watermark;

-**Capacity**, the amount of bits that can be hidden through watermark algorithms

-**Computational complexity**, the global watermark system should be computationally efficient.

A specific parameter for live audio watermark system is ***the delay*** between the audio data and the watermark. The temporal masking effects suggest that this delay should be close to 20ms in order to ensure the imperceptibility of the watermark.

Next we focus on how the concept of live watermarking can be applied to audio integrity checks in forensic applications. A typical audio recording scene (see Fig. 1) implies a recording device and the persons to be recorded (who might be aware or not that they are recorded).

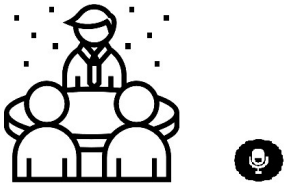

Fig.1. Classic audio surveillance set-up

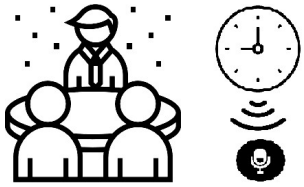

Fig. 2. Audio surveillance set-up with acoustic live watermarking

Further, assuming the audio recording is later used as evidence in a court of law, the person(s) speaking in the recording can claim the audio message is altered, thus the recording is forged. In this situation it can be quite difficult to prove the contrary because the audio integrity has an unpleasant paradigm, i.e. in many cases it is straightforward to prove that an audio recording has been forged, but it is almost impossible to prove 100% that an audio recording has not been forged. This paradigm led us to propose to change the classical recording set-up by adding a watermark in the recording which can be used later to check if the recording (used in a court of law) is original or has been tampered.

An important aspect we need to address is whether the audio forensic benefits best from **offline or live watermarking.** In a court of law any direct interference with the audio recording, including adding an offline watermark, may be seen as an operation which affects the data integrity. Thus, the watermark should be added during the recording of the audio message, as a sound generated by an external audio device, e.g. the clock in Fig. 2. In [10] this acoustic watermark is called sonic watermark.

For live watermarking in audio forensics the performances criteria to be considered are: imperceptibility, security, computation complexity and delay. We should note that in this particular situation, security implies the impossibility for an unauthorized person to be able to extract the watermark and replace it with another one so as to create the impression that a forged recording is authentic. Contrary to expectations, the watermark should not be robust. In fact, we desire for our watermark to be easily affected by malicious attacks, because if we are not able to extract the watermark intact, then we can conclude that the recording has been forged.

As already mentioned, a characteristic of the watermark is its imperceptibility to the human ear. This issue can be addressed in three different ways:

1. The live audio watermark should be generated based on the audio recording in order to benefit from the temporal masking effects; this approach implies several aspects concerning the delay between the watermark and the audio signal and the computational complexity.

2. Use a sound that it can be heard by anyone but it is very familiar to everyone, that it will not be noticed. In this manner we eliminate the problem with the delay between the audio watermark and the audio recording;

3. A combination between the first two solutions: camouflage the audio watermark signal with a sound that can be heard by anyone, however it passes unnoticed.

In our first attempt we used the second solution. We solved the imperceptibility problem by using the 'TIC-TAC' clock sound as a watermark. This very familiar sound solves also the security issue because it has a wide bandwidth (due to its impulse like nature), making it almost impossible for it to be extracted from the audio recording and then replaced with a copy that sustains the forgery. Most likely, noticeable artifacts will be present in the audio recording.

The 'TIC-TAC' sound is in fact a sequence of equally spaced impulse like sounds. Their succession in time is 1 second apart. If an insertion or a deletion of fragments is performed on the audio recording, then the TIC-TAC synchronicity will be destroyed, consequently we would know that the audio integrity of the recording has been damaged. In order to take advantage of this idea, we have to be able to detect the presence of the 'TIC-TAC' sounds with a very good time resolution, so that the synchronicity is not damaged by our own detection algorithm.

The problem with this approach is that the 'TIC-TAC' sound may have a signal to noise ratio – SNR of -20dB, even -30dB, therefore extracting it with 100% accuracy may prove to be quite difficult. Let us mention that the signal in the SNR is the TIC TAC sound, while the noise is the audio/speech recording.

In order to be able to detect the acoustic watermark with a good temporal resolution, in the context of low SNR, we propose to hide a second watermark in the 'TIC-TAC' sound, a known signal based on which an adapted filter may be created. In this manner it is very easy to detect a chirp signal with a bandwidth around 20Hz generated in the spectral regions where the human ear is less sensitive, namely in the low frequency band or around high frequencies. The following section addresses this topic in detail.

3. TIC-TAC LIVE AUDIO WATERMARKING

The TIC-TAC sound has an inner periodicity of 1s (see Fig. 3, I, bottom panel). If we control the generation of its periods by introducing an imperceptible small delay, $\pm\Delta t$ between the TICs and TACs, then we obtain a unique temporal succession and a distance between them varying in the range $[1-2\Delta t;1+2\Delta t]$, as it can be seen in Fig. 3, II, top panel.

When introducing the Δt delay between the TICs and TACs, first we have to ensure that a person would not be able to perceive these variations and second that the chosen variation is much bigger than the temporal resolution of the algorithm used for detecting the 'TIC-TAC' sound. Based on the Δt variations of the 'TIC-TAC' delay we can create a live audio watermark pattern that cannot be noticed by a person being recorded in the room where the 'TIC-TAC' watermark is present.

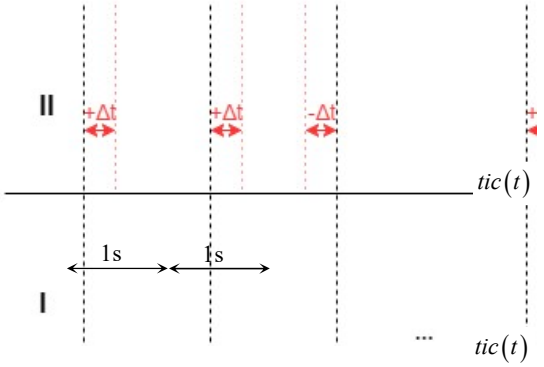

Fig. 3. The 'TIC-TAC' synchronicity mix

The problem with the 'TIC-TAC' watermark is that its SNR, in contrast with the audio that is recorded, may be as low as -30dB, therefore it can be quite difficult to automatically determine the time occurrence of the tics with a good temporal resolution. Due to this aspect we propose to use a second sound, hidden by the 'TIC-TAC' sound, but which is easily traced in low SNR conditions.

Having $s(t)$ the speech/audio sound recorded in a room, the watermark will be denoted with $w(t)$. Since we have no delay between the two signals, the watermarked signal, $s_w(t)$ recorded in the room is obtained as:

$$s_w(t) = s(t) + w(t) . \quad (1)$$

We generated the live audio watermark, $w(t)$ based on the following equation:

$$w(t) = tic(t) + x(t) , \quad (2)$$

where $tic(t)$ is a 'TIC-TAC' sound with various $\pm\Delta t$ delays and $x(t)$ is a chirp signal, masked by the $tic(t)$ sound. This idea was inspired by the signals usually used in radar applications.

The recorded signal $s_w(t)$ is then passed through a matched filter $h(t)$ with $x(t)$, in order to determine the temporal synchronicity of the live audio watermark.

The matched filter is defined as the conjugated time-reversed version of $x(t)$,

$$h(t) = x^*(T - t) , \quad (3)$$

where T – is the duration of the chirp signal.

The output of the adapted filter will be

$$s_a(t) = (s_w * h)(t) , \quad (4)$$

where $(*)$ denotes the convolution operation.

In Fig. 4a (top panel) one can see the recorded signal with the 'TIC-TAC' live audio watermark, $s_w(t)$, while in Fig. 4b (bottom panel) it is illustrated the output of an adapted filter with the live audio watermark, $s_a(t)$. The dashed lines mark the time moments of the watermark's occurrence, which are synchronized with the highest local maxima in the waveform of $s_a(t)$.

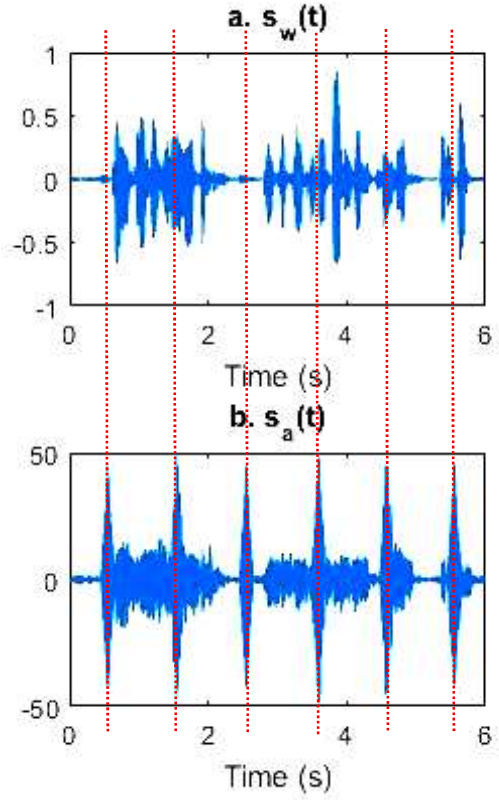

Fig. 4. a. Example of $s_w(t)$, b. Example of $s_a(t)$

We propose as live watermark a ten seconds long audio signal, with a pattern of ten $\pm\Delta t$ delays which will repeat as long as the signal is recorded. The live watermark will be generated by a custom designed wall clock so that it will not to be noticed by the persons being recorded in the room. The wall clock is designed to have a specific TIC-TAC pattern, any pattern out of the 2^{10} possible patterns. In fact, this means that we can use 2^{10} different clocks, placed in as many rooms. Another function of the clock is to monitor the sound level in the room in order to adjust the level of the watermark as low as possible so that the person in the room will not notice it, however the level should be high enough so that the SNR of the watermark is greater than -30dB. Regarding the chirp signal $x(t)$, its bandwidth is between 10Hz and 30Hz so that only an authorized person will know the adapted filter, which will function as a key for extracting the watermark.

4. INTEGRITY CHECK BASED ON TIC-TAC LIVE AUDIO WATERMARKING

In the previous section we have seen how the 'TIC-TAC' watermark can be incorporated and how it can be extracted using an adapted filter.

Based on the signal, $s_a(t)$ we determine the 'TIC-TAC' watermark occurrence moments by searching for the temporal coordinates of the local maxima in the signal's waveform. These

time coordinates are stored as a discrete signal $t_m(n)$ in order to be later compared with the watermark pattern for integrity check.

The data integrity is checked by comparing the signal $t_m(n)$ with the presumed (and known) incorporated 'TIC-TAC' delay pattern.

The integrity check algorithm is based on the Pearson correlation coefficient computed between the 1st derivative of the signal $t_m(n)$ and the pattern signal $p(n)$, as in (5). The calculus is performed frame-wise, with k denoting the frame index.

$$r_{tp}(k) = \frac{\sum_{n=1}^{10}(t_{2m}^{k}(n) - \overline{t_{2m}^{k}})(p(n) - \overline{p})}{\sqrt{\sum_{n=1}^{10}(t_{2m}^{k}(n) - \overline{t_{2m}^{k}})^2}\sqrt{\sum_{n=1}^{10}(p(n) - \overline{p})^2}} \quad (5)$$

The signal $t_{2m}^{k}(n)$ is the k^{th} analysis frame from the first derivative of $t_m(n)$, while $\overline{(\cdot)}$ denotes the mean. The integrity check algorithm is schematically presented in Fig. 5.

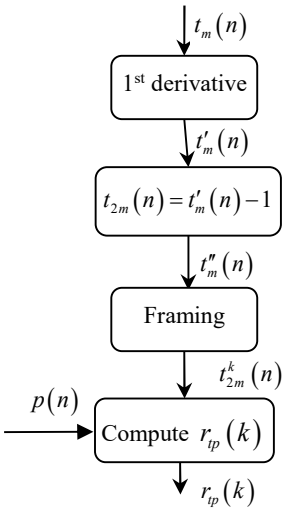

Fig. 5. Schematic of the integrity check algorithm based on the Pearson correlation coefficient.

We proposed to test the integrity check performances by analyzing 100 clean recordings and 100 edited recordings based on the 'TIC-TAC' watermark integrity check. We use the following statistical metrics to quantify the performance of our algorithm:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (6)$$

The performance results for both clean and tampered recordings are synthetized in Fig. 6. One can observe that the 'TIC-TAC' may be a powerful tool for determining the integrity of a recording with respect to cut and paste tampering. The algorithm has a 100% precision in forgery detection if the length of the cuts is greater than 5ms.

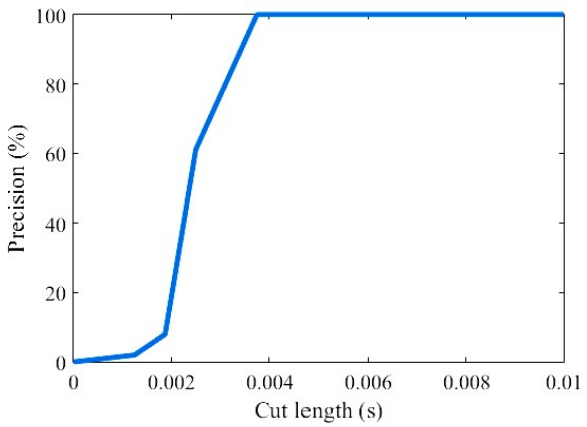

Fig. 6. The precision for finding cut tampering based on the length of the cut

Also, if the recordings are clean, not affected by any cut/paste tampering, the false alarm rate is 0%, making the proposed solution a very reliable tool that can be used in audio forensics.

5. CONCLUSIONS

In this paper we introduced a novel live audio watermarking technique used in the context of audio forensics. Our approach provides a reliable solution for authenticating audio recordings, by deliberately inserting a control signal (the watermark) in the audio recording. Let us stress that the insertion process is performed simultaneously with the speech/audio recording, therefore the integrity of the latter signal is not altered. As watermark we proposed to use a TIC-TAC sound like to benefit from its familiarity due to its occurrence in almost any daily activity.

The originality of our solution lies not only in the chosen watermark, but also in the manner in which we generated the watermark itself. Namely, we introduced a random sequence of delays in the succession of the tics, which serves as a pattern for authenticating the recording. Moreover, to make the solution robust to noise we used a second control signal, the chirp signal, which acts as a key for extracting the pattern. The chirp signal is inaudible, as it is masked by the TIC-TAC signal. Using an adapted filter we proved how we can extract the watermark signal (hence the pattern of delays) from the mixture of signals: the speech/audio signal, the TIC-TAC signal and the chirp signal. Next we provided an algorithm for integrity check, based on the Pearson correlation coefficient. Finally we evaluated the precision of the algorithm for a reduced set of edited and unedited signals, recorded in the TIC-TAC context, and we obtained very good results. Specifically, for insertion or extraction of segments greater than 5ms we obtain 100% precision. Moreover, the false alarm rate is 0% for the unedited signals which strongly encourages the usage of this method in a court of law.

In the foreseeable future we intend to extend the testing procedures with perceptual listening test and real life scenarios.

6. REFERENCES

[1] C. Grigoraș, „Digital audio recording analysis: the Electric Network Frequency (ENF) criterion," The International Journal of Speech Language and the Law, vol. 12, nr. 1, pp. 63-76, 2005.

[2] G. Richard, T. Virtanen, J. P. Bello, N. Ono și H. Glotin, „Introduction to the Special Section on Sound Scene and Event Analysis," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, nr. 6, pp. 1169 - 1171, 2017.

[3] S. Ikram și H. Malik, „Microphone Identification Using Higher-Order Statistics," în 46th International Conference: Audio Forensics , Denver, 2012.

[4] D. P. N. Rodrıguez, J. A. Apolinario și L. W. P. Biscainho, „Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," IEEE Trans. Inf. Forensics Security, vol. 5, nr. 3, p. 534–543, 2010.

[5] A. Cooper, „Detecting butt-spliced edits in forensic digital audio recordings," în Proc. Audio Eng. Soc. 39th Conf., Audio Forensics: Practices and Challenges, Hillerod, 2010.

[6] D. Luo, R. Yang și J. Huang, „Detecting double compressed AMR audio using deep learning," în IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, 2014.

[7] H. Zhao și H. Malik, „AUDIO FORENSICS USING ACOUSTIC ENVIRONMENT TRACES," în IEEE Statistical Signal Processing Workshop (SSP), Ann Arbor, 2012.

[8] A. Ciobanu, T. Culda, C. Negrescu și D. Stanomir, „Analysis of reverberation time blind estimation used in audio forensics," în 11th International Symposium on Electronics and Telecommunications (ISETC), Timisoara, 2014.

[9] W.-H. Chuang, R. Garg și M. Wu, „Anti-forensics and countermeasures of electrical network frequency analysis," IEEE Trans. Inf. Forensics Security, vol. 8, nr. 12, pp. 2073-2086, 2013.

[10] R. Tachibana, „Sonic Watermarking," EURASIP Journal on Applied Signal Processing, vol. 13, p. 1955–1964, 2004.

[11] G. Hua, J. Huang, Y. Q. Shi, J. Goh și V. L. Thing, „Twenty years of digital audio watermarking — a comprehensive review," Elsevier Signal Processing, nr. 128, pp. 222-242, 2016.