DOI: 10.24412/2413-2527-2022-331-61-67

Analysis Method of the Stability of the Combined Labeling of Digital Audio Signals

Grand PhD A. A. Kornienko, PhD M. V. Gofman, PhD S. V. Kornienko
Emperor Alexander I St. Petersburg State Transport University
Saint Petersburg, Russia
kaa.pgups@ya.ru, offmail3000@mail.ru, sv.diass99@ya.ru

Abstract. **The article analyzes the stability of the method of combined marking of digital audio signals. The following are the main characteristics of labelling quality and stability: auditory transparency (inaudible) marker, lossy compression resistance and resistance to stegoanalysis, airway transmission immunity. The article contains quantitative estimates of indicators for the specified characteristics which received analytically from simulations and field experiments. To assess auditory transparency (inaudible) of acoustic artifacts that appear as a result of the use of the marker analytical method is used, which is based on the signal-marker ratio metric. Lossy compression stability is applied to MP3 transformations under different compression modes. Resistance to stegoanalysis is evaluated using a universal method, which uses the ratio of number of units to zeros in the least significant digitized audio signal reference grades. Evaluation of audio immunity over the air audio channel by correct/incorrect rate of marker rectoration based on the results of field experiments which were implemented in laboratory settings.**

Keywords: **steganography, digital audio signal, combined labeling, quality and sustainability evaluation, air audio channel, immunity transmission.**

Introduction

One of the main problems with steganographic protection of audio information and copyright is the problem of quality assurance and sustainability of digital labeling to various transition, to stegoanalysis, to the action of noise and liability in audio channel. This is particularly important when the required sustainability in the marking of digital audio signals which are embedded in audiostego systems with multiple (spatial) inputs and multiple (spatial) outputs is ensured. Such audio stegosystem have several acoustic dynamics (transmitters) and several microphones (receivers) and allows to implement a label in digital audio signal settings to spatial, frequency and time areas, it means that is to do a combination labeling. The spatial multichannel of such audio stegosystem also makes it possible to use methods MIMO-technologies to achieve the required immunity.

Quality analysis of labeling digital audio signals is primarily related to evaluation of auditory transparency (inaudible) or audibility of acoustic artifacts resulting from the implementation of the label that defines thresholds for the insertion force of a label.

Audibility (non-audibility) evaluation methods of acoustic artifacts are divided into analytical methods which use different metrics or indicators [1–3], automatic methods [4–8], based on the use of various psychoacoustic models of human hearing organs, manual methods of hearing evaluation of markers [9], most of which are test adaptations that used to assess the quality of audio information.

Evaluation of the sustainability of digital audio signal labeling involves the use of methods to assess sustainability to digital transformation, to stegoanalysis, acoustic noise/liability resistance (immunity) in the channel, for example, in the air acoustic channel, sustainability to other impacts.

Assessing the resistance of labeling methods to transformations is reduced to assessing the resistance to various types of attacks on labelled audio information and audio signals. The article [10] focusses on the sustainability of digital data labelling. Attacks on digital data labelling systems can be divided into the following: active attacks, passive attacks, introduction attacks, collusion attacks, simple attacks, ambiguity attacks, removal attacks, synchronization attacks and other attacks. Simple attacks are the most commonly used type of attacks. The article [11] describes Stirmark software that allows you to perform simple attacks. Simple attacks on digital audio signal, for example, include a lossy compression attack.

An overview of digital audio stegoanalysis methods is presented in article [12]. Methods of stegoanalysis of digital audio signals can be divided into two classes. A class of methods aimed at stegoanalysis of audio signals presented in uncompressed format (for example, FLAC or WAV), and a class of methods aimed at stegoanalysis of audio signals presented in compressed format (for example, MP3 or AAC). In turn, the first class of methods can be divided into two subclasses. The first subclass includes methods focused on predefined implementation algorithms — these are the so-called «targeted» methods of stegoanalysis. The second subclass includes methods in which it is either not expected to have knowledge about the implementation algorithm in advance, or a certain set of possible or assumed implementation algorithms are known — these are the so-called «universal» methods of stegoanalysis. The class of universal methods can also be divided into two subclasses. The first subclass is methods in which the decision on the presence of a marker is made only on the basis of the properties of the stegoanalysed audio signal. The second subclass is methods in which the solution is made on the basis of comparing the properties of a stegoanalysed audio signal with the properties of some other audio signal (usually obtained from a stegoanalysed signal), regarding which it is precisely known whether it is a stegoaudio signal or just an audio signal. Among the universal methods of stegoanalysis, the most common method based on extraction and analysis of the least significant bit, or LSB stegoanalysis [13].

Evaluation of the noise immunity of marked audio signals when transmitted over an air channel can be obtained based on the results of full-scale experiments and simulation results. Full-scale experiments on the transmission of a marked audio signal can be broadcasting of marked audio signals using acoustic speakers while recording of the broadcast audio signal with an acoustic microphone to subsequently check whether the marker has been preserved in the recorded audio signal recorded by the microphone or not. Simulation modelling is based on the use of some model of air audio channel, which simulates the impact on the marked audio signal with subsequent assessment of the safety of the marker in the digital audio signal received at the output of the channel.

This article analyses and evaluates the main indicators of quality and stability of the method of combined digital marking of audio signals proposed and developed in articles [14–16] for audio stegosystems with multiple input and multiple output. The quality of the method is characterized by hearing transparency (inaudible). To assess the auditory transparency (inaudibleness) of acoustic artefacts that appear as a result of the introduction of a marker, the article uses an analytical method based on the signal-to-marker relationship metric.

Resistance to lossy compression attack is carried out in relation to MP3 transformations by simulation. Stegoanalysis resistance is assessed using the universal least significant bit method (LSB method), which uses the ratio of the number of units to zeros in the least significant bits of the marked digital audio signal. The assessment of noise immunity of the transmission of marked audio signals through an air audio channel is carried out on the basis of the results of full-scale experiments performed in the laboratory.

A software and hardware model were developed for simulation modelling of digital transformations (attacks). MATLAB was used as a software development environment for the model.

100 musical compositions included in the collection «100 Greatest 00s: The Best Songs from the Decade» were used as the original digital audio signals for labelling. Each of the original digital audio signals had the following parameters: bit depth — 16 bits/countdown, sampling rate — 44 100 countdown/sec, two audio tracks in each digital audio signal (stereo signals). Thus, the digital speed of all audio signals used was the same and equal: 16 × 44 100 × 2 = 1 411 200 bit/s.

Analysis and results of auditory transparency assessment

The auditory transparency assessment of the marker was carried out using the analytical method of checking the audibility of artefacts from the introduction of the marker. The verification method was based on the value of the signal-to-marker ratio

$$\mathrm{SWR} = 10\lg\frac{\sum_n x_n^2}{\sum_n (x_n - y_n)^2},$$

where x_n and y_n — are the references of the original and labelled digital audio signals, respectively. The evaluated labelling method refers to a type of methods that mask the marker in the frequency region of the Fourier spectrum. For such methods, a value of 20 dB is usually used as a threshold for the signal-to-marker ratio. Thus, if for the analyzed marked digital audio signal, the value of this ratio is not less than this threshold, it is highly likely that the human ear will not hear acoustic artefacts resulting from the introduction of the marker.

As sequences $\boldsymbol{\alpha}_i = (\alpha_1\, \alpha_2 \cdots \alpha_{N_\alpha})$ and $\boldsymbol{\beta}_i = (\beta_1\, \beta_2 \cdots \beta_{N_\beta})$, used in the process of converting information according to the marking methods described in the works [14–16], Kasami and Gold sequences were used, respectively.

Sequences $\boldsymbol{\gamma}_i = (\gamma_1\, \gamma_2 \cdots \gamma_{N_\gamma})$, taking into account the similarity of the amplitude spectra of adjacent blocks, length N_{block}, readings of a separate audio track of the marked audio signal, were designed in accordance with the rule

$$\gamma_i = \varphi_i \otimes \mathrm{RZcode}\,,$$

where $\otimes$ — is Kronecker's product, φ_i — is Kasami sequences of length 15, and $\mathrm{RZcode} = (1 \quad -1)$.

Hearing transparency was assessed with different sets of parameters $\{N_\alpha, N_\beta, N_\gamma, N_{block}\}$, but with a fixed embedding force of 0.1.

- At $N_\alpha = 1023, N_\beta = 31, N_\gamma = 30, N_{block} = 64$ and the force of embedding 0.1 of the marker element, the average value of the signal-to-marker ratio calculated for all 100 analysed musical compositions was 24.80661 dB, with the dispersion of this ratio equal to 2.382539 dB.
- At $N_\alpha = 1023, N_\beta = 63, N_\gamma = 30, N_{block} = 128$ and embedding force 0.1, the average value of the signal-to-marker ratio calculated from all 100 analysed musical compositions was 23.00086 dB, with the dispersion of this ratio equal to 1.074242 dB.
- At $N_\alpha = 1023, N_\beta = 127, N_\gamma = 30, N_{block} = 256$ and embedding force 0.1, the average value of the signal-to-marker ratio calculated for all 100 analysed musical compositions was 21.70835 dB, with the dispersion of this ratio equal to 0.439844 dB.
- At $N_\alpha = 1023, N_\beta = 511, N_\gamma = 30, N_{block} = 1024$ and embedding force 0.1, the average value of the signal-to-marker ratio calculated for all 100 musical compositions analysed was 20.68482 dB, with the dispersion of this ratio equal to 0.0446 dB.

Based on these results, it can be concluded that at fixed values N_α and N_γ, the increase in N_β and N_{block} leads to a decrease in the average value and variance of the signal-to-marker ratio. Thus, the analysed marking method provides auditory transparency to the marker when assessing audibility using an analytical method based on the value of the signal-to-marker ratio, when a value of 20 dB is used as a threshold.

Loss resilience assessment analyses and results

The evaluation of the resistance of the analyzed labelling method to a lossy compression attack using MP3 conversion was performed for different MP3 conversion modes and different bit compression speeds. Resistance to MP3 conversion in stereo, mono and joint stereo modes was checked. Combinations of modes and compression speeds were selected so that the original sampling rate of the compressed digital audio signal was maintained.

The MP3 attack procedure consisted of the following sequence of actions. Initially, the original digital audio signal stored in WAVE format was labelled using the developed labelling method. Further, the marked digital audio signal was subjected to an MP3 attack by compression in a certain mode and at a certain bit rate of compression. After that, the MP3 file was decompressed into a WAV file. Finally, it was checked whether the digital audio signal received at the MP3 decoder output retained the embedded token or not. SOUND FORGE Pro 14.0 Suite was used to perform an MP3 attack.

MP3 compression parameters in stereo mode were as follows: compression quality — Fastest encode; bit rate — 96 kbit/s; sampling rate — 44 100 Hz. MP3 decompression parameters: format — PCM (uncompressed); sampling rate — 44 100 Hz; number of bits to count — 16; number of audio tracks — 2 (stereo). The used bit rate of 96 kbit/s is the minimum allowed in this audio processing program at a sampling rate of 44 100 Hz in stereo mode. After the MP3 compression attack in stereo mode, all markers were detected without errors and the information bits encoded in them were restored without errors.

MP3 compression parameters in mono mode were as follows: compression quality — Fastest encode; bit rate — 48 kbit/s; sampling rate — 44 100 Hz. MP3 compression parameters: format — PCM (uncompressed); sampling rate — 44 100 Hz; number of bits to count — 16; number of audio tracks — 1 (mono). The used bit speed of 48 kbit/s is the minimum allowable audio processing program at a sampling rate of 44 100 Hz in mono mode. After an MP3 compression attack in mono mode, all markers without errors were detected and the information bits encoded in them were restored without errors.

MP3 compression parameters in joint stereo mode were as follows: compression quality — Fastest encode; bit rate — 96 kbit/s; sampling rate — 44 100 Hz; Mid/side stereo and Intensity stereo are used. MP3 decompression parameters: format — PCM (uncompressed); sampling rate — 44 100 Hz; number of bits to count — 16, number of audio tracks — 2 (stereo). The used bit speed of 48 kbit/s is the minimum allowable audio processing program at a sampling rate of 44 100 Hz in mono mode. The used bit rate of 96 kbit/s is the minimum allowed in this audio processing program at a sampling rate of 44 100 Hz in joint stereo mode. The joint stereo mode leads to the fact that after the reverse conversion to WAVE Stereo format, each of the two audio tracks will include not only for the most part (since after compression with loss some data is irretrievably lost) the original audio track, but also partially another audio track, which means that two markers can be found in each of the audio tracks. Usually, the lower the bit rate, the more audio tracks «mix» together and the more likely it is to detect a marker from the neighboring audio track in the analyzed audio track. After an MP3 attack in joint stereo mode at a bit speed of 96 kbit/s after analysing 200 audio tracks (100 stereo audio signals) in which each of the two markers was searched (thus, 400 checks were performed: two markers were searched in each of the 200 audio tracks after decompression, due to the possible introduction of a marker from the:

- 14 times out of 400 times a marker brought from the neighboring audio track was not found (this indicator is affected by bit speed);
- 0 times out of 400 times the original marker was not found on the corresponding audio track;
- 11 times out of 400 times there was an error in establishing synchronization with the introduced token (i. e. the introduced token was detected, but its beginning was not correctly defined);
- 3 times out of 400 times there was an error in establishing synchronization with the original token (i. e. the original marker of the corresponding track was detected, but its beginning was not correctly determined).

However, when the bit rate was increased to 320 kbit/s, no synchronization errors were found with the original token — all original tokens were detected and all information bits were correctly restored. It should be noted that at a speed of 320 kbit/s, the mixing of audio tracks as a result of compression was so small that markers introduced from neighboring audio tracks were not found in audio tracks after decompression.

Analysis and results of stegoanalysis resistance evaluation

Using the computer model of the audio stegosystem, the stability of the analyzed marking method to the least significant bit method (LSB method) was assessed. This method of stegoanalysis is a universal method, as it does not require knowledge of the marking method. The method of stegoanalysis of marked digital audio signals using the least significant bit method is based on the assumption that if the ratio of the number of units to zeros in binary bits of the analyzed binary plane of the digital signal reference values exceeds some threshold, it is likely that there is a hidden message in this bit plane. Usually, the threshold is 1.1. It is taken as a threshold. All 100 original digital audio signals have been labelled. All bit planes of reference of marked digital audio signals were analyzed. The sustainability of the developed marking method to stegoanalysis was assessed under different sets of parameters $\{N_\alpha, N_\beta, N_\gamma, N_{block}\}$:

- At $N_\alpha = 1023, N_\beta = 31, N_\gamma = 30, N_{block} = 64$ and embedding force 0.1, the average value of the ratio of units to zeros in binary bits along all 16 bit planes (musical compositions were stored in files with 16 bit depth of reference values) of all 100 analyzed musical compositions was 0.993058, with the dispersion of this ratio equal to 0.00080572.
- At $N_\alpha = 1023, N_\beta = 63, N_\gamma = 30, N_{block} = 128$ and embedding force 0.1, the average value of the ratio of units to zeros in binary bits along all 16-bit planes (musical compositions were stored in files with 16-bit depth of reference values) of all 100 analyzed musical compositions was 0.997205415, with the dispersion of this ratio equal to 0.000459767.

Thus, the developed marking method is resistant to LSB stegoanalysis, providing average values of the number of units to zero ratios for all bit planes are much closer to 1 than to threshold 1.1, while the variance of the average value is less than 10^{-3}.

Analysis and results of the assessment of noise immunity of the transmission of marked audio signals through the air audio channel

The evaluation of the stability of the developed method of marking digital audio signals to the interference of the air audio channel was carried out based on the results of full-scale experiments using a software and hardware laboratory unit. The software part of the laboratory installation was realized in the MATLAB environment. The hardware of the laboratory unit was as follows: the AKG Lyra 4-capsule microphone in Tight Stereo mode was used as an acoustic microphone, and two pairs of Edifier R1280DB speakers were used as acoustic speakers.

Four Kasami sequences of the same length $N_\alpha = 1023$ were used as vectors $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\alpha}_3$ and $\boldsymbol{\alpha}_4$ and four vectors $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$ as Gold sequences of the same length $N_\beta = 63$ were used. When constructing vectors $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, $\boldsymbol{\gamma}_3$ and $\boldsymbol{\gamma}_4$ as vectors $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$, $\boldsymbol{\varphi}_3$ and $\boldsymbol{\varphi}_4$ four different Kasami sequences of the same length 15 were used, and the vector $(1 \quad -1)$ was used as a RZcode word; as a result, $N_\gamma = 30$. The size of the reference block was equal to $N_{block} = 2(N_\beta + 1) = 128$. At such values of marker parameters and sampling rate of F_s = 44 100 Hz, the transmission of one marker takes

$$\frac{N_\alpha N_{block} N_\gamma}{F_s} \cong 89.1 \text{ s} .$$

The value for the amount (determining the size of the analysis window) was taken to 8 093.

Two types of indoor field experiments were carried out to assess the resistance to noise and interference of the air audio channel. The first type of experiment included transmitting marked stereo audio signals using a pair of speakers and stereo recording of broadcast audio signals using a microphone located in line of sight at a distance of 3 meters. The second type of experiment included the transmission of two marked stereo audio signals with two pairs of speakers and stereo recording of broadcast audio signals using a microphone, also located in line of sight at a distance of 3 meters. The second type of experiment simulated a possible situation in practice when there is a significant extraneous acoustic noise during broadcasting.

During the first type of experiment, each of the 100 source stereo audio signals was labelled. Two markers were introduced into each of the two audio tracks of the original digital audio signal. If the duration of the audio signal was not enough to introduce two markers, the audio signal was artificially extended by cyclic repetition from the beginning. At the same time, markers in one audio track were separated from each other in time by a protective interval of 10 % of their own duration. Transmission (broadcasting) through the air audio channel of different marked stereo audio signals was carried out separately. Thus, 400 markers were transmitted through the air audio channel. Speaker amplifiers were set to 4 out of 50 divisions — this gain was minimal, at which the broadcast audio signal was barely audible. The results of the first type of experiment showed that the developed marking method is resistant to the influences of an air audio channel when using Edifier R1280DB speakers and an AKG Lyra microphone located at a distance of 3 meters in line of sight. Thus, all 400 markers were found and the information bits in them encoded were correctly restored.

For the second type of experiment, 50 digital stereo audio signals were randomly selected from 100 original stereo audio signals, which were labelled. Two markers were introduced into each of the two audio tracks. If the duration of the audio signal was not enough to introduce two markers, the audio signal was artificially extended by cyclic repetition from the beginning. At the same time, markers in one audio track were separated from each other in time by a protective interval of 10 % of their own duration. After marking, 25 pairs were randomly formed from 50 marked audio signals. Each stereo audio signal from the pair was broadcast to the air audio channel simultaneously with another stereo audio signal from the same pair using four speakers. This broadcasting of quadro-audio signals, consisting of four audio tracks, simulated the situation of significant extraneous acoustic noise. Thus, 200 markers were transmitted through the air audio channel. Speaker amplifiers were installed on 10 out of 50 divisions. The second type of field experiments showed that the developed marking method is quite resistant to the influences of the air audio channel when using Edifier R1280DB speakers and the AKG Lyra microphone, located at a distance of 3 meters in line of sight. As a result of the transfer, only three markers out of 200 could not detect and restore information.

Conclusion

The article analyzed and evaluated various indicators of quality and stability of the combined method of marking digital audio signals, developed in articles [14–16]. The results of hearing transparency analysis showed that the developed marking method provides a signal-to-marker ratio of at least 20 dB, which is sufficient to prevent acoustic artefacts that inevitably appear as a result of the introduction of markers into digital audio signal. The results of the lossy compression resistance assessment showed that in the case of MP3 compression in mono and stereo modes, the embedded marker can also be detected and the information transferred by it can be restored. When MP3 compression is used in joint stereo mode, then, due to the mixing of stereo audio tracks during compression, synchronization errors may occur, but when the bit rate is 320 kbit/s, then there are no synchronization errors. The results of the lossy compression resistance assessment showed that in the case of MP3 compression in mono and stereo modes, the embedded marker can also be detected and the information transferred by it can be restored. When MP3 compression is used in joint stereo mode, then, due to the mixing of stereo audio tracks during compression, synchronization errors may occur, but when the bit rate is 320 kbit/s, then there are no synchronization errors. The results of the assessment of stegoanalysis resistance using the universal least significant bit method (LSB method) showed that changes in digital audio signals are such that the one-to-zero ratios in individual bit planes of the marked digital audio signal do not exceed threshold 1.1, which in the research article [13] is designated as an indicator of the presence of a marker in digital audio signal. The results of the experimental noise resistance assessment of the air audio channel showed that it is possible to steadily transmit marked stereo audio signals. Under the specified conditions of the experiment, it was possible to detect and restore information sequences of 197 markers out of 200.

References

1. Hu Y., Loizou P. C. Evaluation of Objective Quality Measures for Speech Enhancement, *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, Vol. 16, No. 1, Pp. 229–238. DOI: 10.1109/TASL.2007.911054.

2. Özer H., Sankur B., Memon N., Avcıbaş İ. Detection of Audio Covert Channels Using Statistical Footprints of Hidden Messages, *Digital Signal Processing*, 2006, Vol. 16, Is. 4, Pp. 389–401. DOI: 10.1016/j.dsp.2005.12.001.

3. Kutter M., Petitcolas F. A. P. Fair Benchmark for Image Watermarking Systems. In: *Wong P. W., Delp III E. J. (eds.) Proceedings of the SPIE/IS&T Electronic Imaging 1999. Security and Watermarking of Multimedia Contents, San Jose, CA, USA, January 23–29, 1999. Proceedings of SPIE*, 1999, Vol. 3657, Pp. 226–239. DOI: 10.1117/12.344672.

4. GOST R 54711-2011. Zvukovoe veshchanie tsifrovoe. Kodirovanie signalov tsifrovogo veshchaniya s sokrashcheniem izbytochnosti dlya peredachi po tsifrovym kanalam svyazi. MPEG-1 chast III (MPEG-1 audio) [Digital sound broadcasting. Coding of sound broadcasting signals with redundancy reduction for transfer on digital communication

channels. MPEG-1 part III (MPEG-1 audio)]. Effective from July 01, 2013. Moscow, StandartInform Publishing House, 2014, 172 p. (In Russian)

5. GOST R 54712-2011. Zvukovoe veshchanie tsifrovoe. Kodirovanie signalov tsifrovogo veshchaniya s sokrashcheniem izbytochnosti dlya peredachi po tsifrovym kanalam svyazi. MPEG-2 chast III (MPEG-2 audio) [Digital sound broadcasting. Coding of sound broadcasting signals with redundancy reduction for transfer on digital communication channels. MPEG-2 part III (MPEG-2 audio)]. Effective from July 01, 2013. Moscow, StandartInform Publishing House, 2014, 109 p. (In Russian)

6. Recommendation ITU-R BS.1196-8. Audio coding for digital broadcasting. Approved in 14.10.2019. International Telecommunication Union, 2019, 43 p. Available at: http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1196-8-201910-I!!PDF-E.pdf (accessed 08 Nov 2021).

7. Recommendation ITU-R BS.1387-1. Method for objective measurements of perceived audio quality. Approved in 24.11.2001. International Telecommunication Union, 2001, 100 p. Available at: http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1387-1-200111-I!!PDF-E.pdf (accessed 08 Nov 2021).

8. Tiede T., Treurniet W., Bitto R., et al. PEAQ — The ITU Standard for Objective Measurement of Perceived Audio Quality, *Journal of the Audio Engineering Society*, 2000, Vol. 48, No. 1, Pp. 3–29.

9. Hua G., Huang G., Shi Y. Q., et al. Twenty Years of Digital Audio Watermarking — A Comprehensive Review, *Signal Processing*, 2016, Vol. 128, Pp. 222–242.
DOI: 10.1016/j.sigpro.2016.04.005.

10. Agarwal N., Singh A. K., Singh P. K. Survey of Robust and Imperceptible Watermarking, *Multimedia Tools and Applications*, 2019, Vol. 78, Is. 7, Pp. 8603–8633.
DOI: 10.1007/s11042-018-7128-5.

11. Steinebach M., Petitcolas F. A. P., Raynal F., et al. StirMark Benchmark: Audio Watermarking Attacks, *Information Technology: Coding and Computing (ITCC 2001): Proceedings of the II International Conference, Las Vegas, NV, USA, April 02–04, 2001*. Institute of Electrical and Electronics Engineers, 2001, Pp. 49–54.
DOI: 10.1109/ITCC.2001.918764.

12. Ghasemzadeh H., Kayvanrad M. H. Comprehensive Review of Audio Steganalysis Methods, *IET Signal Processing*, 2018, Vol. 12, Is. 6, Pp. 673–687.
DOI: 10.1049/iet-spr.2016.0651.

13. Dittmann J., Hesse D. Network Based Intrusion Detection to Detect Steganographic Communication Channels: On the Example of Audio Data, *Proceedings of the IEEE 6th Workshop on Multimedia Signal Processing, Siena, Italy, September 29–October 01, 2014*. Institute of Electrical and Electronics Engineers, 2004, Pp. 343–346.
DOI: 10.1109/MMSP.2004.1436563.

14. Gofman M. V. Razrabotka modeli mnogokanalnoy audiostegosistemy na osnove markirovaniya tsifrovykh audiosignalov [Development of a Model of a Multi-Channel Audio Stegosystem Based on Digital Audio Watermarking], *Sovremennaya nauka: aktualnye problemy teorii i praktiki. Seriya: Estestvennye i Tekhnicheskie Nauki [Modern Science: Actual Problems of Theory and Practice. Series: Natural and Technical Sciences]*, 2021, No. 6, Pp. 78–83.
DOI: 10.37882/2223–2966.2021.06.11. (In Russian)

15. Gofman M. V. Obnaruzhenie markera v tsifrovom audiosignale avtorizovannym poluchatelem [Detection of Marker in Digital Audio Signal by Authorized Recipient], *Sovremennaya nauka: aktualnye problemy teorii i praktiki. Seriya: Estestvennye i Tekhnicheskie Nauki [Modern Science: Actual Problems of Theory and Practice. Series: Natural and Technical Sciences]*, 2021, No. 2, Pp. 45–50.
DOI: 10.37882/2223-2966.2021.02.09. (In Russian)

16. Gofman M. V., Kornienko A. A., Mironchikov E. T., Nikitin A. B. Tsifrovoe markirovanie audiosignalov dlya robastnoy skrytoy akusticheskoy svyazi cherez vozdushnyy audiokanal [Digital Watermarking of Audio Signals for Robust Hidden Audio Communication via Air Audio Channel], *Trudy SPIIRAN [SPIIRAS Proceedings]*, 2017, Is. 6 (55), Pp. 185–215. DOI: 10.15622/sp.55.8. (In Russian)

DOI: 10.24412/2413-2527-2022-331-61-67

Анализ устойчивости метода комбинированного маркирования цифровых аудиосигналов

д.т.н. А. А. Корниенко, к.т.н. М. В. Гофман, к.т.н. С. В. Корниенко
Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
kaa.pgups@ya.ru, offmail3000@mail.ru, sv.diass99@ya.ru

Аннотация. **В статье анализируется устойчивость метода комбинированного маркирования цифровых аудиосигналов к преобразованиям и помехам, применяемого в аудиостегосистемах с множественным (пространственным) входом и множественным (пространственным) выходом. В качестве основных характеристик устойчивости используются следующие: слуховая транспарентность (неслышимость маркера), устойчивость маркирования к сжатию с потерями (к MP3-преобразованиям), помехоустойчивость передачи через воздушный аудиоканал. Статья содержит количественные оценки показателей для указанных характеристик, полученные в результате натурных экспериментов и имитационного моделирования.**

Ключевые слова: **стеганография, цифровой аудиосигнал, комбинированное маркирование, оценка качества и устойчивости, воздушный аудиоканал, помехоустойчивость передачи.**

Литература

1. Hu, Y. Evaluation of Objective Quality Measures for Speech Enhancement / Y. Hu, P. C. Loizou // IEEE Transactions on Audio, Speech, and Language Processing. 2008. Vol. 16, No. 1. Pp. 229–238. DOI: 10.1109/TASL.2007.911054.

2. Detection of Audio Covert Channels Using Statistical Footprints of Hidden Messages / H. Özer, B. Sankur, N. Memon, İ. Avcıbaş // Digital Signal Processing. 2006. Vol. 16, Is. 4. Pp. 389–401. DOI: 10.1016/j.dsp.2005.12.001.

3. Kutter, M. Fair Benchmark for Image Watermarking Systems / M. Kutter, F. A. P. Petitcolas // Proceedings of the SPIE/IS&T Electronic Imaging 1999. Security and Watermarking of Multimedia Contents (San Jose, CA, USA, 23–29 January 1999) / P. W. Wong, E. J. Delp III (eds.). Proceedings of SPIE. 1999. Vol. 3657. Pp. 226–239. DOI: 10.1117/12.344672.

4. ГОСТ Р 54711-2011. Звуковое вещание цифровое. Кодирование сигналов цифрового вещания с сокращением избыточности для передачи по цифровым каналам связи. MPEG-1 часть III (MPEG-1 audio) = Digital sound broadcasting. Coding of sound broadcasting signals with redundancy reduction for transfer on digital communication channels. MPEG-1 part III (MPEG-1 audio): национальный стандарт Российской Федерации: утвержден и введен в действие приказом Федерального агентства по техническому регулированию и метрологии от 13 декабря 2011 г. № 872-ст: дата введения 2013–07–01. — Москва: Стандартинформ, 2014. — 172 с.

5. ГОСТ Р 54712-2011. Звуковое вещание цифровое. Кодирование сигналов цифрового вещания с сокращением избыточности для передачи по цифровым каналам связи. MPEG-2 часть III (MPEG-2 audio) = Digital sound broadcasting. Coding of sound broadcasting signals with redundancy reduction for transfer on digital communication channels. MPEG-2 part III (MPEG-2 audio): национальный стандарт Российской Федерации: утвержден и введен в действие приказом Федерального агентства по техническому регулированию и метрологии от 13 декабря 2011 г. № 873-ст: дата введения 2013–07–01. — Москва: Стандартинформ, 2014. — 109 с.

6. Recommendation ITU-R BS.1196-8. Audio coding for digital broadcasting. Approved in 14.10.2019. — International Telecommunication Union, 2019. — 43 p. URL: http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1196-8-201910-I!!PDF-E.pdf (дата обращения 08.11.2021).

7. Recommendation ITU-R BS.1387-1. Method for objective measurements of perceived audio quality. Approved in 24.11.2001. — International Telecommunication Union, 2001. — 100 p. URL: http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1387-1-200111-I!!PDF-E.pdf (дата обращения 08.11.2021).

8. PEAQ — The ITU Standard for Objective Measurement of Perceived Audio Quality / T. Tiede, W. Treurniet, R. Bitto, [et al.] // Journal of the Audio Engineering Society. 2000. Vol. 48, No. 1. Pp. 3–29.

9. Twenty Years of Digital Audio Watermarking — A Comprehensive Review / G. Hua, G. Huang, Y. Q. Shi, [et al.] // Signal Processing. 2016. Vol. 128. Pp. 222–242. DOI: 10.1016/j.sigpro.2016.04.005.

10. Agarwal, N. Survey of Robust and Imperceptible Watermarking / N. Agarwal, A. K. Singh, P. K. Singh // Multimedia Tools and Applications. 2019. Vol. 78, Is. 7. Pp. 8603–8633. DOI: 10.1007/s11042-018-7128-5.

11. StirMark Benchmark: Audio Watermarking Attacks / M. Steinebach, F. A. P. Petitcolas, F. Raynal, [et al.] // Information Technology: Coding and Computing (ITCC 2001): Proceedings of the II International Conference (Las Vegas, NV, USA, 02–04 April 2001). — Institute of Electrical and Electronics Engineers, 2001. — Pp. 49–54. DOI: 10.1109/ITCC.2001.918764.

12. Ghasemzadeh, H. Comprehensive Review of Audio Steganalysis Methods / H. Ghasemzadeh, M. H. Kayvanrad // IET Signal Processing. 2018. Vol. 12, Is. 6. Pp. 673–687. DOI: 10.1049/iet-spr.2016.0651.

13. Dittmann, J. Network Based Intrusion Detection to Detect Steganographic Communication Channels: On the Example of Audio Data / J. Dittmann, D. Hesse // Proceedings of the IEEE 6th Workshop on Multimedia Signal Processing (Siena, Italy, 29 September–01 October 2014). — Institute of Electrical and Electronics Engineers, 2004. — Pp. 343–346. DOI: 10.1109/MMSP.2004.1436563.

14. Гофман, М. В. Разработка модели многоканальной аудиостегосистемы на основе маркирования цифровых аудиосигналов // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки. 2021. № 6. С. 78–83.
DOI: 10.37882/2223-2966.2021.06.11.

15. Гофман, М. В. Обнаружение маркера в цифровом аудиосигнале авторизованным получателем // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки. 2021. № 2. С. 45–50. DOI: 10.37882/2223-2966.2021.02.09.

16. Цифровое маркирование аудиосигналов для робастной скрытой акустической связи через воздушный аудиоканал / М. В. Гофман, А. А. Корниенко, Е. Т. Мирончиков, А. Б. Никитин // Труды СПИИРАН. 2017. Вып. 6 (55). С. 185–215. DOI: 10.15622/sp.55.8.