https://doi.org/10.30764/1819-2785-2019-14-4-125-136

Forensic Investigation of MP3 Audio Recordings

Alexander G. Boyarov[1], Ivan S. Siparov[2]

[1] The Russian Federal Centre of Forensic Science of the Ministry of Justice of the Russian Federation, Moscow 109028, Russia

[2] North-West Regional Centre of Forensic Science of the Ministry of Justice of the Russian Federation, Saint-Petersburg 191014, Russia

Abstract. Special aspects of MP3-recordings technical investigation are addressed. The following features of formation and research of MP3 phonograms are explained: traces of MP3 coding in time and spectral domain, special aspects of MP3-files structure analysis, detection methods of re-coding of MP3-recordings, methods of group identification of MP3-recorders and MP3-codecs.

MP3 coding leaves certain traces of its usage. Due to the psychoacoustic model inaudible spectral components are deleted from the signal spectrum. Traces of psychoacoustic codecs usage are also clearly seen via dynamic spectrogram as rectangular areas of zero spectral amplitude. The methods discussed in this paper enable the investigating expert to detect the exact position of the MP3 frame in the signal by its properties even without any information from the file header. This method reveals the coding itself, multiple coding and also audio editing by the investigation of the periodicity of the extracted frames' positions.

MP3 file format specifies the structure of the frame header providing a perfect instrument to detect any periodicity of any peculiarities of MP3 frames. The tool based on this approach reveals MP3 frames disorder caused by editing in the "digital" domain – manual deletion of audio information using HEX editor.

Keywords: *forensic video and audio examination, digital audio forensics, MP3-encoding, detection of double MP3 compression, tampering detection, audio authenticity analysis, MP3-frames headers, frame shift offsets, SC-fragment, MP3 file frame allocation map, PN-formula*

For citation: Boyarov A.G., Siparov I.S. Forensic Investigation of MP3 Audio Recordings. *Theory and Practice of Forensic Science*. 2019. Vol. 14. No. 4. P. 125–136. https://doi.org/10.30764/1819-2785-2019-14-4-125-136

Криминалистическое исследование фонограмм формата MP3

А.Г. Бояров[1], И.С. Сипаров[2]

[1] Федеральное бюджетное учреждение Российский федеральный центр судебной экспертизы при Министерстве юстиции Российской Федерации, Москва, 109028, Россия

[2] Федеральное бюджетное учреждение Северо-Западный региональный центр судебной экспертизы Министерства юстиции Российской Федерации, Санкт-Петербург 191014, Россия

Аннотация. Рассмотрены специальные аспекты технического исследования фонограмм в формате MP3. Раскрыты особенности формирования и исследования MP3-фонограмм: следы кодирования MP3 во временной и спектральной областях, особенности анализа структуры MP3-файлов, методы обнаружения перекодирования MP3-записей, методы групповой идентификации MP3-рекордеров и MP3-кодеков.

Кодирование MP3 оставляет определенные следы его использования. Из-за психоакустической модели неразличимые спектральные компоненты удаляются из спектра сигнала. Следы использования психоакустических кодеков также хорошо видны на динамической спектрограмме в виде прямоугольных областей с близкой к нулевой энергией спектра. Представленные в статье методы позволяют определить точное положение фрейма MP3 в сигнале по его свойствам. Предлагаемые методы выявляют кодирование, множественное кодирование, а также редактирование аудио без перекодирования – путем исследования периодичности обнаруживаемых по карте размещения фреймов.

Формат файла MP3 определяет структуру заголовка фрейма, обеспечивая обнаружение любой периодичности всех особенностей MP3-фреймов. Инструмент, основанный на этом подходе, выявляет нарушения порядка в MP3-фреймах, обусловленные редактированием в «цифровой» области – ручным удалением аудиоинформации с помощью HEX-редактора.

Ключевые слова: *экспертиза видео- и звукозаписей, экспертиза цифровых фонограмм, MP3-кодирование, поиск монтажа, заголовки MP3-фреймов, анализ сдвига границ фреймов, НС-фрагмент, карта размещения фреймов MP3-файлов, PN-формула*

Для цитирования: Бояров А.Г., Сипаров И.С. Судебно-экспертное исследование фонограмм формата MP3 // Теория и практика судебной экспертизы. 2019. Т. 14. № 4. С. 125–136. (Англ.). https://doi.org/10.30764/1819-2785-2019-14-4-125-136

Introduction

One of the key demands to audio recording as a case evidence is its authenticity. Authenticity analysis of speech recordings is a typical task of forensic audio examination. Even negative result of recordings' integrity gives the basis for further investigation.

Nowadays the majority of speech recordings are created by the means of digital audio recorders and the volume of recorded media is only increasing just as in all the other aspects of human life. Audio file format differs from device to device and supports different types of coding and compressing algorithms.

One of the most widespread audio file formats is MP3. Popularity of this format is historically based on psychoacoustics approach – the fact that the codec should not tamper with audio components that in any way (corresponding to particular sound environment) cannot be heard by a human's ear (frequency and time masking effects) and, thus, can be skipped during compression. So, ceteris paribus, the listener cannot hear the difference between original and compressed data.

The branch of audio codecs (audio coding algorithms) that today is called MP3 was developed at the end of XX century by MPEG GROUP. This branch includes codecs standards MPEG 1[1], MPEG 2[2] and MPEG 2.5 with three Layers: I, II and III. The most widespread is Layer III – MP3 format and MP3 file extension. Due to openness of engine source producers of digital recorders proceed to develop a number of modifications, and even under new names.

This article is dedicated to revealing the traces of MP3 coding that can be helpful during authenticity analysis of evidential audio, recorded (or received by an investigator) in MP3 file format. In other words of those that can testify the integrity and authenticity of the audio. The article is based on real cases studies and represents the development of methods reported in [1].

Nowadays there are many scientific articles on forensic audio authentication. Review articles like [2–4] and also papers dedicated to particular methods of analysis are present. For example there is a set of articles dedicated to ENF analysis like [5–9] and focused on background noise research like [10]. Also, acoustic environment analysis [11] and microphone identification [12] are well studied. Papers like [13, 14] are dedicated to investigation of MP3, being focused on MDCT-analysis.

In this article we assume that the recording device which was used to create the audio should also be submitted for examination, though in practice it is often not the case.

This article captures those traces and aspects of MP3 recordings authenticity analysis that were detected and extracted during several years of expert practice and aims to create the source for further practical work.

Research

During the research traces of MP3 coding were divided into six groups (corresponding to the influence on different aspects of audio recording and coding).

1. Text content of MP3 files headers.
2. Traces of MP3 coding in time domain.
3. Traces of MP3 coding in frequency domain.
4. MPEG-frames header analysis.
5. Frame allocation map of MP3 files.
6. Stereo modes of MP3 recordings.
7. Frame Offsets Check Method.

The following sections describe each of these seven types of traces.

Text content of MP3 files headers

The information listed in MP3 file header can be and must be used for the means of authenticity analysis because it contains the metadata filled by the device coder and other features. The match of file header structure and content of header fields for evidential recording and testing recording is an important fact that should be mentioned in the expert's report.

The list of header fields and possibilities of their content are well known [2]. The most interesting are the following.

Block "TENC" (encoder) in metadata of ID3v2 [2, 3] format is reserved for codec information.

In some of audio recordings created by Sony voice recorders "TENC" block can con-

[1] ISO/IEC 11172-3:1993 Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio.

[2] ISO/IEC 13818-3:1998 Information technology – Generic coding of moving pictures and associated audio information – Part 3: Audio.

tain “SONY IC RECORDER MP3...”; Panasonic: “PanaICR”, the devices’ model and ID and so on.

These should be checked and compared to those in test recordings.

“LAME” codec (typically used on PCs) “Info” block (sometimes substitute first frame) can contain: codec version (e.g., “LAME3.98”); number of frames; hash sum; coding parameters; etc.

Besides, information about LAME codec (name and version) can be located inside data field of MP3 files.

Revealing of such circumstances helps to prove the match of evidence audio file header structure and content to those of the test files recorded with the device used to create the evidence. The difference of these features testify that the recording was re-coded and/or edited.

Traces of MP3 coding in time domain (detecting traces of repeated or “double” MP3 coding)

In case when original MP3 recording was converted and edited (by means of any software sound editor) it has to be converted back to correspond to original file format – this leads to “repeated” or “double” conversion.

Previous researches in this area showed good results only when the final bitrate was greater than the initial. If the expert has access to the recording device, the practice shows it is possible to detect double conversion in time domain.

Usage of some of MP3 codecs leads to appearance of a short fragment (usually less than 1 second) that does not reflect the real audio events. The amplitude of this fragment (*Starting Coder fragment – SC-fragment*) is quite low (sometimes it even contains consequent zero samples) so its border is clearly seen thanks to the rise of amplitude and appearance of actual audio environment in the signal.

Individual features of *Starting Coder fragment* (duration, dynamic amplitude characteristics) are different for different codecs and even for different modes of the same codec. Every next coding adds new *SC-fragment* to the signal. Thus, having a test recording from the investigated device, the expert can compare the *SC-fragment* features in test recording and in the evidential recording.

Following experiment reveals traces of double coding: Microsoft PCM file was converted to MP3, decoded back to WAV and converted to MP3 again. The result of such experiment with *LAME 3.98.4* codec is shown on Figure 1.

The duration of *SC-fragment* for *LAME 3.98.4* codec (11025 Hz sampling rate) is 1105 samples. Right border of *SC-fragment* is clearly seen and can be easily detected during accurate visual waveform analysis (Figure 2).

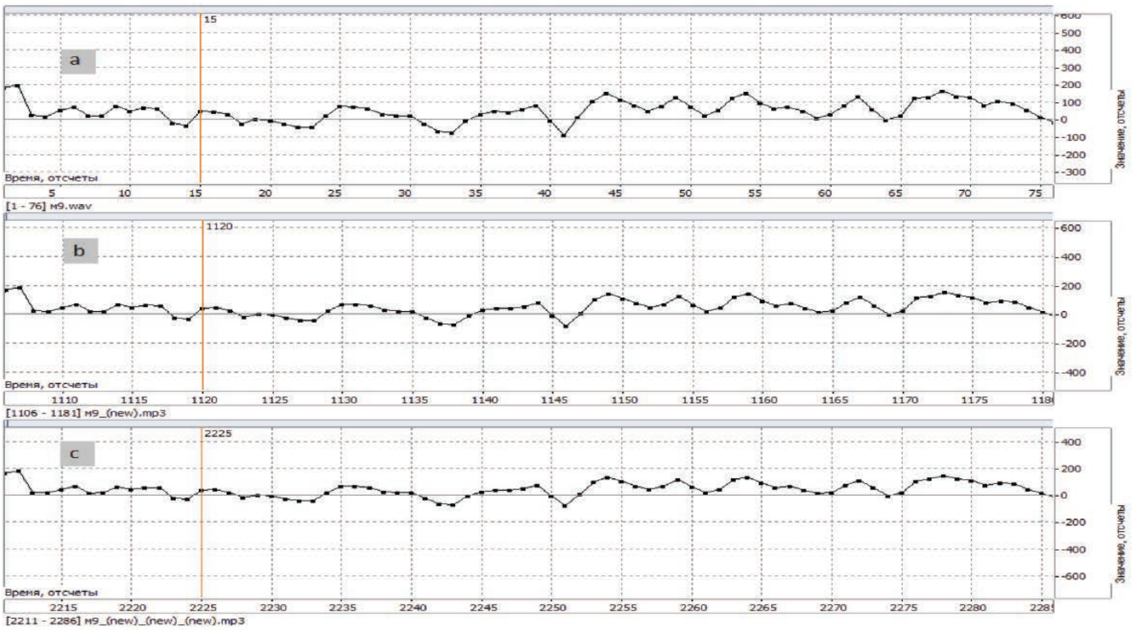

Fig. 1. Waveforms of the same pattern in the signal: *a* – before MP3 coding; *b* – after first MP3 coding; *c* – after second MP3 coding. The time coordinate (in samples) of the pattern (marked with cursor) is increasing due to adding of SC-fragments during MP3 coding

Рис. 1. *Осциллограммы одного и того же фрагмента сигнала: a – до кодирования, b – после однократного кодирования и c – после повторного кодирования. Время в отсчетах образца (отмеченного курсором) увеличивается на длительность начального фрагмента, добавляемого при кодировании*

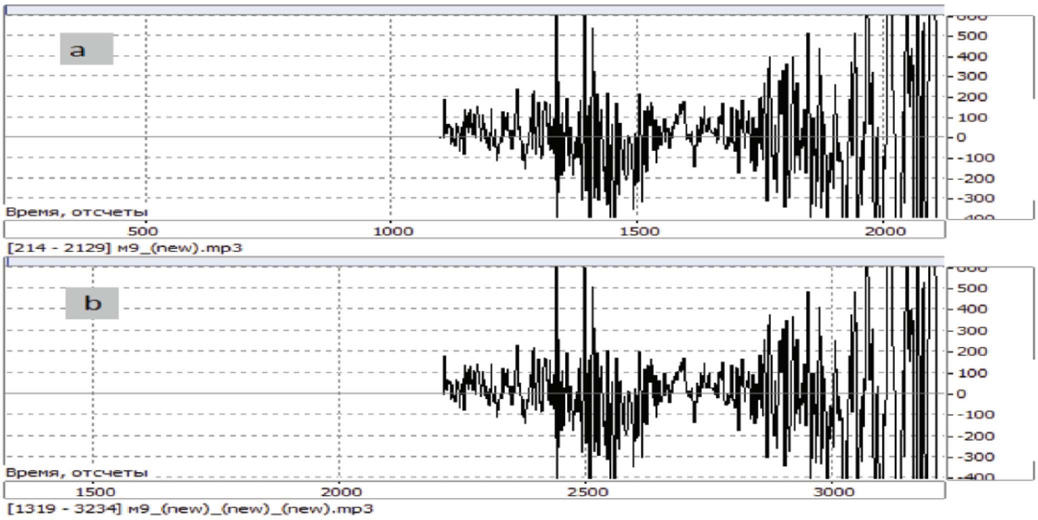

Fig. 2. Waveforms (time in samples) of SC-fragment: *a* – after first MP3 coding; *b* – after second MP3 coding. Double coded signal contains two SC-fragments

Рис. 2. *Осциллограммы начального фрагмента сигнала: а – после однократного и b – после повторного MP3-кодирования. Дважды кодированный сигнал содержит два начальных фрагмента.*

SC-fragment of some digital recorders (suppoting MP3 file format) are very individual (Figure 3). This fact can be used to identify the recorder and, comparing to test recordings, prove the authenticity of the recording in question.

Thus, *SC-fragment* can be used for the means of authenticity analysis: the duration and amplitude dynamic shape of *SC-fragment* in test and evidential recordings must match to prove the integrity and authenticity of the recorded audio information. Double duration of *SC-fragment* corresponds to double coding which should have understandable reasons.

Traces of MP3 coding in spectral domain

Average spectra of investigated recordings are traditionally used to reveal traces of resampling or re-coding of the signal. In addition, the average spectrum carries information about frequency response of the recording channel and can be used for device identification. The analysis of these characteristics in case of MP3 recordings should consider the traces of the codec in spectral domain because it contains features similar to those mentioned above.

Figure 4 represents average spectrum of the recordings after it was coded with MPEG1 Layer II. The spectrum has two roll-offs (at 8 000 Hz and 16 000 Hz) which are typical for resampling traces. But in this case these characteristics have origin in codecs' frequency response limitations. The dynamic spectrogram (Figure 5) finely resolute different fragments of the signal that were coded with different frequency response limitations, though the recording is continuous. These fragments give different contribution to the average spectrum.

Average spectra of different fragments of the *same* recording are represented in Figure 6. Blue curve represents the fragment where the speech signal is rather quiet (so the codec leaves spectral components onto 15 000 Hz). Red curve corresponds to the fragment where the speech is comparatively loud (so "there is

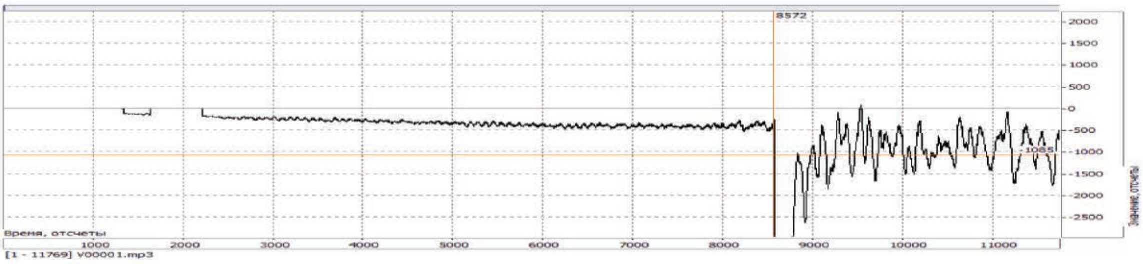

Fig. 3. Samsung YP-U3 PPIM SC-fragment waveform. The right border of SC-fragment (marked with cursor) is clearly seen. SC-fragment waveform shape is unique

Рис. 3. *Осциллограмма участка сигнала, содержащего начальный фрагмент медиаплеера Samsung YP-U3 PPIM. Начальный фрагмент имеет характерную форму для данной модели устройства записи*

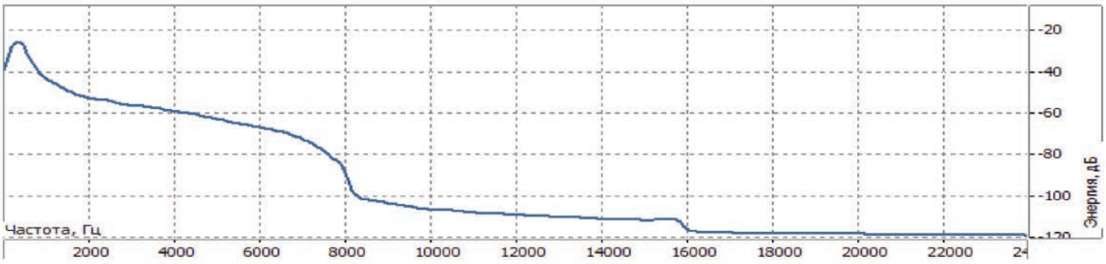

Fig. 4. Average spectrum of the signal coded with MPEG1 Layer II. Roll-offs at 8 and 16 kHz are clearly seen
Рис. 4. *Интегральный спектр сигнала после кодирования кодеком MPEG1. Хорошо видны «завалы» в спектре на частотах 8 и 16 кГц*

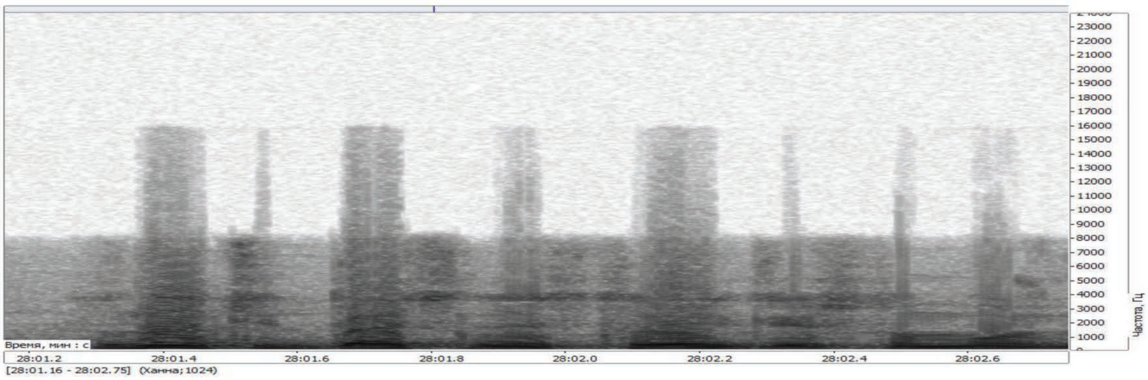

Fig. 5. Dynamic spectrogram of the signal coded with MPEG1 Layer II.
Different fragments have different frequency range
Рис. 5. *Динамическая спектрограмма сигнала после кодирования MPEG1 Layer II.
Разные фрагменты сигнала имеют разный частотный диапазон*

no reason" for the codec to save upper components). Thus, the codec with the same settings can demonstrate different frequency range limitations depending on the energy of different components of the signal.

The particular characteristics of the codec implemented in the recording device should be revealed during device examination. Test recordings must differ not only by technical characteristics (bitrate, sampling rate, "quality") but also by audio environment of test recordings to represent the codec behavior in different acoustics. Test recordings by audio environment (noises, geometry of the room, etc.) must match the investigated recording situation to represent particularities of the codecs' impact. Only after establishing the codecs' influence in spectral domain the investigation can be proceeded to "traditional" types of analysis and interpretation.

MPEG frames header analysis

Each MPEG frame contains 4 bytes header. The table below represents MPEG frame header fields with description and commentaries. There are three frame header fields that can give additional information for comparative analysis of evidential and test recordings.

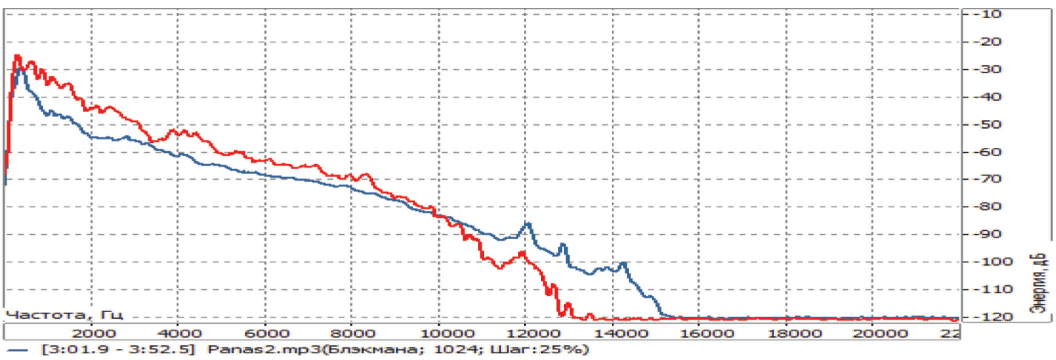

Fig. 6. Average spectra of different fragments of the same recording. Frequency ranges are different, though the recording is not interrupted
Рис. 6. *Интегральные спектры различных фрагментов одной фонограммы имеют разный частотный диапазон*

Table. MPEG frame header structure
Таблица. Структура заголовка фрейма формата MPEG

#	Size, bit	Description
1	11	*syncword* ‘1111 1111 111’
2	2	*ID* – one bit to indicate the ID of the algorithm. Equals ‘1’ for MPEG audio, ‘0’ is reserved
3	2	*Layer* – to indicate which layer is used, according to the following: ‘11’ Layer I ‘10’ Layer II ‘01’ Layer III ‘00’ reserved
4	1	*protection_bit* – to indicate whether redundancy has been added in the audio bitstream to facilitate error detection and concealment. Equals ‘1’ if no redundancy has been added, ‘0’ if redundancy has been added
5	4	*bit_rate_index* – indicates the bitrate. The all zero value indicates the ‘free format’ condition, in which a fixed bitrate which does not need to be in the list can be used. Fixed means that a frame contains either N or N+1 slots, depending on the value of the padding bit. The bit_rate_index is an index to a table, which is different for the different layers.
6	2	*sampling_frequency* – indicates the sampling frequency, according to the following table. ‘00’ 44.1 kHz ‘01’ 48 kHz ‘10’ 32 kHz ‘11’ reserved
7	1	*padding_bit* – if this bit equals ‘1’ the frame contains an additional slot to adjust the mean bitrate to the sampling frequency, otherwise this bit will be ‘0’. Padding is only necessary with a sampling frequency of 44.1 kHz. For MPEG Layer III this bit is used for sampling frequencies 11025, 22050 and 44100 Hz (can be ‘0’ or ‘1’). For all other sampling frequencies it is ‘0’.
8	1	*private_bit* – bit for private use. Is not used for coding. *Can be used for authenticity analysis.*
9	2	*mode* – indicates the mode according to the following: ‘00’ stereo ‘01’ joint_stereo (intensity_stereo and/or ms_stereo) ‘10’ dual_channel ‘11’ single_channel In Layer I and II the joint_stereo mode is intensity_stereo, in Layer III it is intensity_stereo and/or ms_stereo. *Important aspect for authenticity analysis.*
10	2	*mode_extension* – these bits are used in joint_stereo mode. In Layer I and II they indicate which subbands are in intensity_stereo. All other subbands are coded in stereo. ‘00’ subbands 4–31 in intensity_stereo, bound==4 ‘01’ subbands 8–31 in intensity_stereo, bound==8 ‘10’ subbands 12–31 in intensity_stereo, bound==12 ‘11’ subbands 16–31 in intensity_stereo, bound==16 In Layer III they indicate which type of joint stereo coding method is applied. The frequency ranges over which the intensity_stereo and ms_stereo modes are applied. intensity_stereo ms_stereo ‘00’ off off ‘01’ on off ‘10’ off on ‘11’ on on
11	1	*copyright* – if this bit equals ‘0’ there is no copyright on the coded bitstream, ‘1’ means copyright protected. *Can be used for authenticity analysis*.
12	1	*original/home* – this bit equals ‘0’ if the bitstream is a copy (usually set by audio software apps), ‘1’ if it is an original (usually set by recording devices). (Can be changed by third party programs.) *Can be used for authenticity analysis.*
13	2	*emphasis* – indicates the type of de-emphasis that shall be used (very rare use). ‘00’ no emphasis ‘01’ 50/15 microsec. emphasis ‘10’ reserved ‘11’ CCITT J.17

There are three frame header fields containing information that do not influence the coding itself and store the information about the file: *Private_Bit*, *Original_Bit* and *Copyright_Bit*.

The values of these bits are set by codec depending on the recording settings and should be used for files comparison. There are 8 possible combinations that should match for both evidential and test recordings. It should be taken in consideration, of course, that those fields can be changed by the means of some software or fixed during re-coding.

Frame allocation map of MP3 files

In MP3 files with sampling rates 11025, 22050 and 44100 Hz the frames with *Padding_Bit* and without it are queued in different sequences (with a period of 49 frames). The sequences can be finely detected over frame allocation map. Figure 7 represents frame allocation maps of MP3 files with 44100 Hz sampling rate and different bitrates (smaller frames are marked dark-grey). Files were conducted with *Mp3Pro* codec.

The periodicity of 49 frames is calculated from the frame size. For 64 kbps bit rate and 44100 Hz sampling rate the frame size is calculated by the following expression:

$$BR(44100{,}64000) = \frac{1152*64\,000}{8*44\,100} = \frac{1152*80}{441} = \frac{128*9*80}{9*49} = 208 + \frac{48}{49} \quad (1)$$

where 1152 – frame shift in samples for MPEG-1 Layer III, 64 000 – bit rate (b/s), 44100 – sampling rate (Hz).

The option of frame size variation on 1 byte was introduced to provide fixed bit rate of audio stream. Thus, in sequence of 49 frames the size of 48 frames will be 209 bytes and the size of 1 frame will be 208 bytes.

For 64 000 bit/s and 11025, 22050 Hz the frame sizes are:

$$BR(11025{,}64000) = \frac{576*64\,000}{8*11\,025} = \frac{64*9*320}{9*49} = \frac{64*320}{49} = 417 + \frac{47}{49}; \quad (2)$$

$$BR(22050{,}64000) = \frac{576*64\,000}{8*22\,050} = \frac{64*9*160}{9*49} = \frac{64*160}{49} = 208 + \frac{48}{49}. \quad (3)$$

The match of frame sizes for 22050 and 44100 Hz sampling rates corresponds to the fact that frame size of MPEG1 is twice bigger than MPEG2.

The frame size and analysis window size can mismatch. In MPEG1 Layer III the frame contains information about two coding windows (“granules” in standards’ description); in MPEG2 and 2.5 Layer III the frame contains information only about one window.

Thus for 44100 Hz sampling rate and 64 kbps bit rate in 49 frames sequence there should be 48 frames with additional byte and 1 frame without additional byte. Figure 8 represents images of frame allocation maps for four different MP3 codec types. Light sectors represent frames with 209 bytes, dark – 208 bytes.

Different MP3 codecs allocate the smaller frame in different position of the sequence. PanalCR puts the smaller frame at the end of the sequence (pos.49), MP3Pro in the middle (pos.25), and LAME 3.98 at the beginning (pos.1). The codec implemented in Samsung

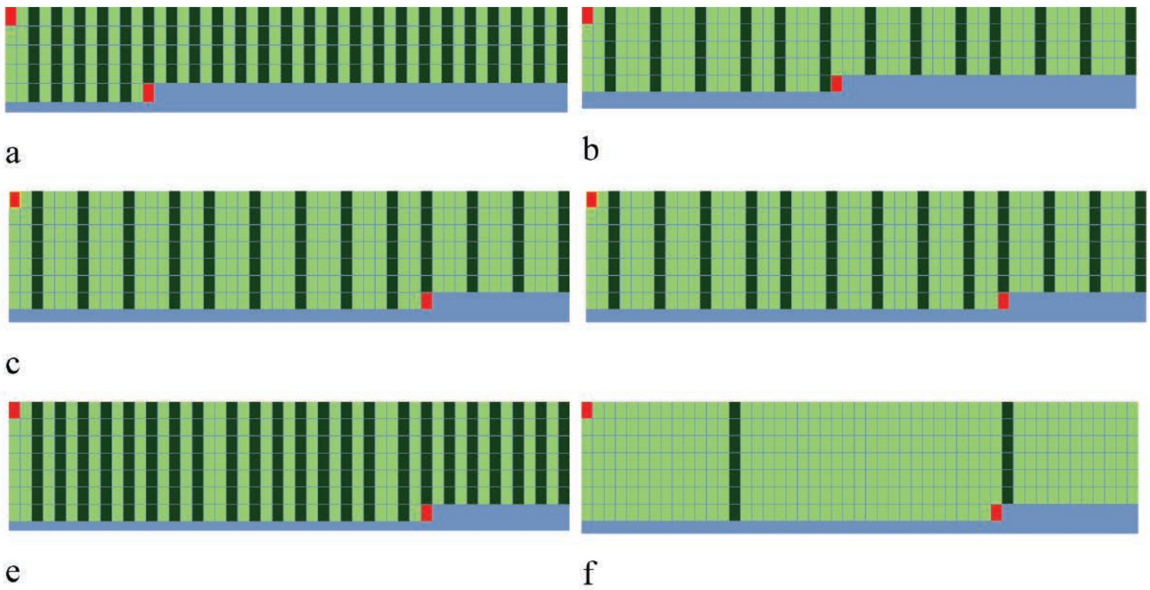

Fig. 7. Frame allocation map for different bitrates:
a – 16 kbps, *b* – 24 kbps, *c* – 48 kbps, *d* – 64 kbps, *e* – 96 kbps, *f* – 128 kbps

Рис. 7. *Размещение блоков в MP3-файле с различными битрейтами:*
a – 16 кбит/с, b – 24 кбит/с, c – 48 кбит/с, d – 64 кбит/с, e – 96 кбит/с, f – 128 кбит/с

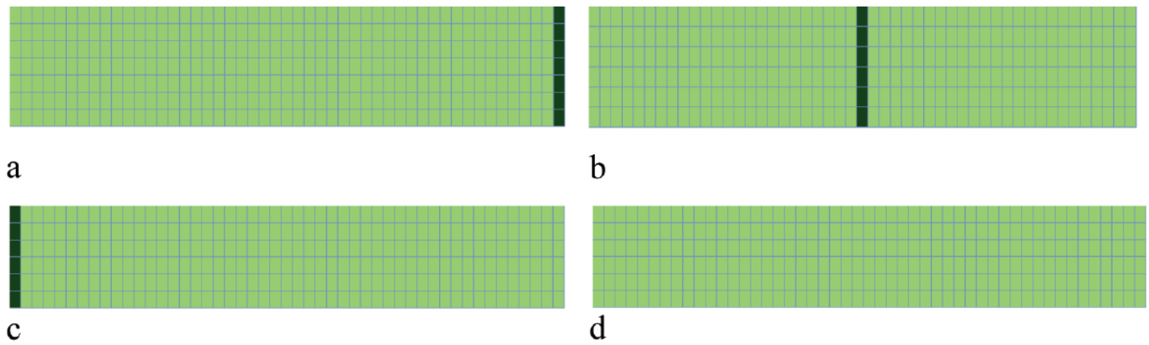

Fig. 8. Frame allocation map for different codecs: *a* – PanalCR codec (voice recorder Panasonic RR-US551), *b* – Mp3Pro codec (software sound editor Adobe Audition 3.0), *c* – LAME 3.98 codec, *d* – codec of digital recorder Samsung YP-U3 PPIM

Рис. 8. *Карта размещения блоков для файла, созданного с использованием различных кодеков: a – кодек PanalCR (цифровой диктофон Panasonic RR-US551), b – кодек Mp3Pro (звуковой редактор Adobe Audition 3.0), c – кодек LAME 3.98, d – кодек, встроенный в плеер Samsung YP-U3 PPIM*

YP-U3 PPIM MP3 player does not use Padding_Bit for precise bit rate evaluation. MP3 codecs usually installed on PC Padding_Bit usage can be chosen in codecs' settings. The frame allocation map for such files will look like represented in Figure 8.

It is important to mention, that *frame allocation map* can represent any other features that are different for frames. So the *frame allocation map* itself is a powerful tool for any frame based codec investigation (e.g. see section 2.6).

The frame sequence can be described with a formula: if N is for frames without Padding_Bit and P – for frames with it, the description for different coders can be expressed as *48P + N* for *PanalCR, 24P + N + 24P* for *Mp3Pro*, N + 48P for *LAME 3.98. PN-formula* for the same codecs must match.

Thus, *PN-formula* and frame allocation map can be used to identify the codecs used to create MP3 recordings. During authenticity analysis of an MP3 recording it is important to compare PN-formulas and frame allocation maps for evidential and test recordings. These features must correspond to the circumstances of the case: *PN-formula* and *frame allocation map* of *LAME 3.98* (used mostly during coding with PC) differs from those implemented in digital recorders.

Stereo modes of MP3 recording. There are four main modes of stereo in MP3 codecs:

- *Mono* – one channel signal;
- *Dual channel* – two channels are coded independently with the same bit rate;
- *Stereo* – signals of two channels are coded with different bit rates, but the sum of bit rates is constant;
- *J-Stereo* – signals of two channels are coded together with two different extensions:
  - *MS Stereo (mide/side);*
  - *Intensity stereo.*

In *MS Stereo* stereosignal is derived from average between channels (up to a factor of (L + R) and differential (up to a factor of (L - R). The bit rate for the "average" signal is greater than for the "differential". So the same general bit rate provides better coding quality (for fragments that has the same phase for left and right channels).

In *Intensity stereo* mode average signal and differences of intensity by ranges are coded only, so the processed data volume decreases. Intensity stereo mode is usually used for low bit rates.

In *J-Stereo* each frame can have its own mode extension due to the parameters of the signal. Figure 9 represents *frame allocation map* where dark regions correspond to frames coded with MS Stereo extension, light frames – without it.

Stereo coding mode can be significant for authenticity analysis. Mismatch of stereo coding mode, listed in the frame header and reflected on the frame allocation map can indicate at least general editing or re-coding. Clarification of stereo coding mode can be achieved by subtraction of channels:

- If left and right channels are equal the difference between them is zero;
- Difference of *Dual channel* or *Stereo* coded signals does not have any peculiarities;
- Difference of *J-Stereo* coded signal contains artifacts, represented in Figure 10.

During MP3 recording investigation the attention should be paid to correspondence of stereo coding mode listed in file header to its features, reflected in the signal and frame allocation maps. For example:

- Listed "*Stereo*" or "*Dual channel*" – traces of *J-Stereo* in channels subtraction;

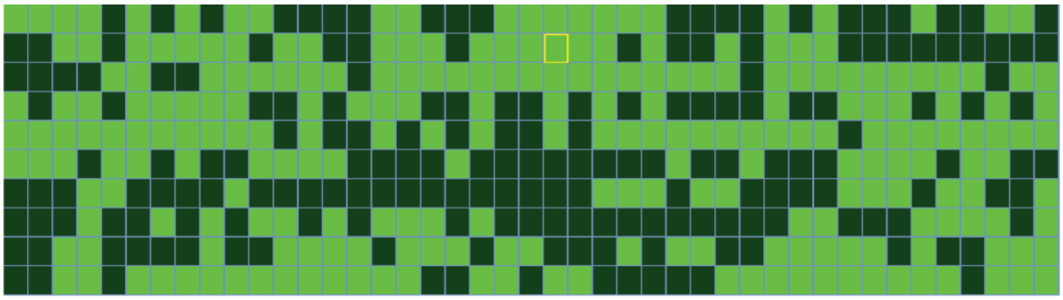

Fig. 9. Frame allocation map for signal coded with J-Stereo. Dark frames correspond to frames coded with MS Stereo, light – without it

Рис. 9*. Карта размещения блоков MP3-файла с использованием режима кодирования J-Stereo*

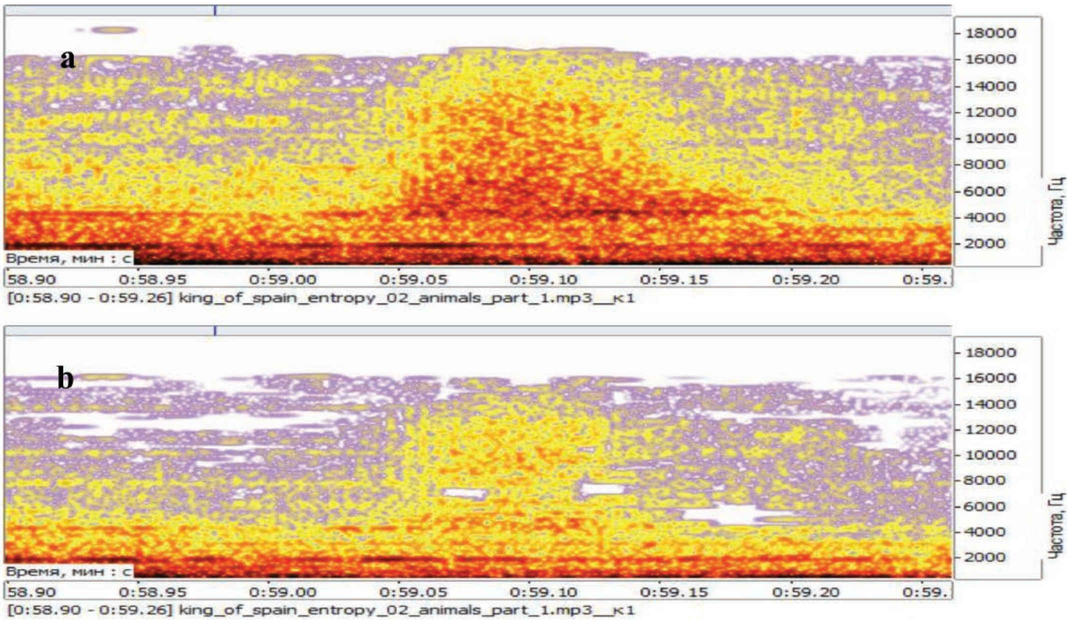

Fig. 10. Dynamic spectrogram of the left channel (*a*) and of the subtraction: left channel minus right channel (*b*)

Рис. 10*. Динамическая спектрограмма сигнала левого канала (a) и фрагмента сигнала левого канала после вычитания из него сигнала правого канала (b)*

• Listed “*Stereo*” or “*Dual channel*” – channels are equal.

These examples prove either re-coding or recorder feature, that should be established during device analysis.

Thus, investigation of stereo MP3 recordings should include analysis of stereo recording mode and its properties that can be traced by signals themselves or its frame allocation map.

Frame Offsets Check Method

Psychoacoustic codecs approach is based on the features of human hearing – the codec deletes spectral components that cannot be heard due to frequency masking. During its workflow the coder calculates signal spectrum on a frame of 1152 samples (frame shift = ½ 1152 = 576 samples); spectrum components that cannot be heard by human’s hearing system are deleted (set to zero); spectral data left is stored into a memory block. These referral frames reconstruct decoded signal (during playback and visualization of the waveform).

If a test frame of 1152 samples is taken and number of zeros in MDCT spectrum is calculated being moved for one sample it provides with a Number-of-Zeros in MDCT time dependence graph.

Sharp peaks in the graph reveal to MP3-frames position; average flat level correspond to test frames’ positions between frames. The behavior of the graph reflects the periodicity of the frames positions and reveals editing points.

Traces of frame coding of the signal (MP3) – the distance of 576 samples correspond to the 50 % frame shift during coding.

Traces of editing: creation of the audio from fragments of different MP3 files. The described procedure allows to calculate the periodical shift between positions of two consequent frames (Figure 11).

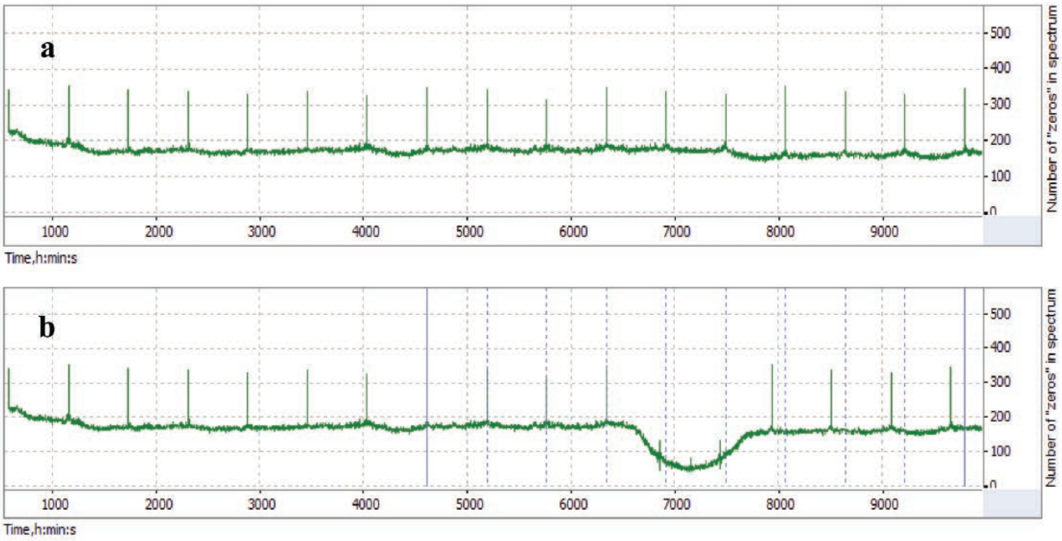

Fig. 11. Number-of-Zeros in MDCT spectra Graph without peaks periodicity broken (*a*), Number-of-Zeros in MDCT spectra Graph with peaks periodicity broken at the position of editing (*b*)

Рис. 11. *Графики количества нулей в спектре МДКП без нарушения периодичности максимумов (а) и с нарушением периодичности максимумов в точке монтажа (b)*

Using this method the expert can solve various tasks:

- Detection of MP3 codec traces;
- Detection of original sampling rate;
- Detection of double coding (with the same or other codec).

And the main thing is the detection of editing traces which were made after signal decoding [115, 126].

Conclusions

Suggested investigation methods enhance technical analysis of MP3 recordings and provide more efficient use of contemporary methods and means for forensic audio examination.

During investigation the expert can determine MP3 files' characteristics and properties and their correspondence to characteristics and properties of test files, created by the device used for recording of the audio evidence.

To ensure the completeness of the examination, following features, properties and/or characteristics should be subjected to careful analysis:

- Text content of header fields of MP3 file;
- *Starting Coder fragment* presence and its duration;
- Average spectrum and changing of the upper border of frequency range depending on audio environment;
- Frame sequence, *PN-formula* and *frame allocation map*;
- Stereo coding mode.

Researches reflected in this paper broaden methodological framework and technical scientific base used for forensic audio authenticity analysis and corrrespond to current state-of-the-art and experience in the field of audio forensics.

REFERENCES

1. Boyarov A.G. Forensic Analysis of MP3 Audio Files. *Theory and Practice of Forensic Science*. 2013. No. 3 (31). P. 70–83. (In Russ.)
2. Ho A.T.S., Li S. (eds). *Handbook of Digital Forensics of Multimedia Data and Devices*. Guildford: Wiley, 2015. 670 p. https://doi.org/10.1002/9781118705773
3. Maher R. Audio Forensic Examination. *IEEE Signal Processing Magazine*. 2009. Vol. 26. No. 2. P. 84–94. https://doi.org/10.1109/msp.2008.931080
4. Gupta S., Cho S., Kuo C-C.J. Current Developments and Future Trends in Audio Authentication. *IEEE Multimedia*. 2012. Vol. 19. No. 1. P. 50–59.

СПИСОК ЛИТЕРАТУРЫ

1. Бояров А.Г. Особенности технического исследования фонограмм формата МР3 // Теория и практика судебной экспертизы. 2013. № 3 (31). С. 70–83.
2. Ho A.T.S., Li S. (eds). Handbook of Digital Forensics of Multimedia Data and Devices. Guildford: Wiley, 2015. 670 p. https://doi.org/10.1002/9781118705773
3. Maher R. Audio Forensic Examination // IEEE Signal Processing Magazine. 2009. Vol. 26. No. 2. P. 84–94. https://doi.org/10.1109/msp.2008.931080
4. Gupta S., Cho S., Kuo C-C.J. Current Developments and Future Trends in Audio Authentication. *IEEE Multimedia*. 2012. Vol. 19. No. 1. P. 50–59.

https://doi.org/10.1109/mmul.2011.74

5. Grigoras C. Applications of ENF Criterion in Forensic Audio, Video, Computer and Telecommunication Analysis. *Forensic Science International*. 2007. Vol. 167. No. 2-3. P. 136–145. https://doi.org/10.1016/j.forsciint.2006.06.033
6. Bessonov A.A., Boyarov A.G., Stepanov M.V. The Electric Network Frequency Based Method of Forensic Audio Analysis. *Theory and Practice of Forensic Science*. 2019. Vol. 14. No. 2. P. 43–50. (In Russ.). https://doi.org/10.30764/1819-2785-2019-14-2-43-50
7. Bykhovsky D., Cohen A. Electrical Network Frequency (ENF) Maximum-Likelihood Estimation Via a Multitone Harmonic Model. *IEEE Transactions on Information Forensics and Security*. 2013. Vol. 8. No. 5. P. 744–753. https://doi.org/10.1109/tifs.2013.2253462
8. Grigoras C., Smith J.M. Digital Imaging: Enhancement and Authentication. In: Siegel J.A., Saukko P.J., Houck M.M. (eds). *Encyclopedia of Forensic Sciences*. Academic Press, 2013. P. 303–314. https://doi.org/10.1016/b978-0-12-382165-2.00127-6
9. Huijbregtse M., Geradts Z. Using the ENF Criterion for Determining the Time of Recording of Short Digital Audio Recordings. In: Geradts Z.J.M.H., Franke K.Y., Veenman C.J. (eds). *Computational Forensics. IWCF 2009. Lecture Notes in Computer Science*. Vol. 5718. Berlin: Springer, 2009. P. 116–124. https://doi.org/10.1007/978-3-642-03521-0_11
10. Pan X., Zhang X., Lyu S. Detecting Splicing in Digital Audios Using Local Noise Level Estimation. *2012 IEEE International Conference on Acoustics*. Speech and Signal Processing. 2012. https://doi.org/10.1109/icassp.2012.6288260
11. Malik H. Acoustic Environment Identification and Its Applications to Audio Forensics. *IEEE Transactions on Information Forensics and Security*. 2013. Vol. 8. No. 11. P. 1827–1837. https://doi.org/10.1109/tifs.2013.2280888
12. Ikram S., Malik H. Digital Audio Forensics Using Background Noise. *2010 IEEE International Conference on Multimedia and Expo*. IEEE, 2010. https://doi.org/10.1109/icme.2010.5582981
13. Yang R., Qu Z., Huang J. Exposing MP3 Audio Forgeries Using Frame Offsets // ACM Transactions on Multimedia Computing, Communications, and Applications. 2012. Vol. 8. No. 2S. P. 1–20. https://doi.org/10.1145/2344436.2344441
14. Liu Q., Sung A.H., Qiao M. Detection of Double MP3 Compression. *Cognitive Computation*. 2010. Vol. 2. No. 4. P. 291–296. https://doi.org/10.1007/s12559-010-9045-4
15. Boyarov A.G. Detection and Analysis of MP3, WMA, OGG AND VORBIS Codecs Traces in Audio Signal. *Theory and Practice of Forensic Science*. 2016. No. 1 (41). P. 78–86. (In Russ.)

https://doi.org/10.1109/mmul.2011.74

5. Grigoras C. Applications of ENF Criterion in Forensic Audio, Video, Computer and Telecommunication Analysis // Forensic Science International. 2007. Vol. 167. No. 2-3. P. 136–145. https://doi.org/10.1016/j.forsciint.2006.06.033
6. Бессонов А.А., Бояров А.Г., Степанов М.В. Метод производства судебной экспертизы фонограмм с использованием данных о частоте электрического тока сети // Теория и практика судебной экспертизы. 2019. Т. 14. № 2. С. 43–50. https://doi.org/10.30764/1819-2785-2019-14-2-43-50
7. Bykhovsky D., Cohen A. Electrical Network Frequency (ENF) Maximum-Likelihood Estimation Via a Multitone Harmonic Model // IEEE Transactions on Information Forensics and Security. 2013. Vol. 8. No. 5. P. 744–753. https://doi.org/10.1109/tifs.2013.2253462
8. Grigoras C., Smith J.M. Digital Imaging: Enhancement and Authentication. In: Siegel J.A., Saukko P.J., Houck M.M. (eds). Encyclopedia of Forensic Sciences. Academic Press, 2013. P. 303–314. https://doi.org/10.1016/b978-0-12-382165-2.00127-6
9. Huijbregtse M., Geradts Z. Using the ENF Criterion for Determining the Time of Recording of Short Digital Audio Recordings. In: Geradts Z.J.M.H., Franke K.Y., Veenman C.J. (eds). Computational Forensics. IWCF 2009. Lecture Notes in Computer Science. Vol. 5718. Berlin: Springer, 2009. P. 116–124. https://doi.org/10.1007/978-3-642-03521-0_11
10. Pan X., Zhang X., Lyu S. Detecting Splicing in Digital Audios Using Local Noise Level Estimation / *2012 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2012. https://doi.org/10.1109/icassp.2012.6288260
11. Malik H. Acoustic Environment Identification and Its Applications to Audio Forensics // IEEE Transactions on Information Forensics and Security. 2013. Vol. 8. No. 11. P. 1827–1837. https://doi.org/10.1109/tifs.2013.2280888
12. Ikram S., Malik H. Digital Audio Forensics Using Background Noise. 2010 IEEE International Conference on Multimedia and Expo. IEEE, 2010. https://doi.org/10.1109/icme.2010.5582981
13. Yang R., Qu Z., Huang J. Exposing MP3 Audio Forgeries Using Frame Offsets // ACM Transactions on Multimedia Computing, Communications, and Applications. 2012. Vol. 8. No. 2S. P. 1–20. https://doi.org/10.1145/2344436.2344441
14. Liu Q., Sung A.H., Qiao M. Detection of Double MP3 Compression // Cognitive Computation. 2010. Vol. 2. No. 4. P. 291–296. https://doi.org/10.1007/s12559-010-9045-4
15. Бояров А.Г. Обнаружение и анализ следов сжатия звуковых сигналов кодеками MP3, AAC, WMA и VORBIS (методические рекомендации для экспертов) // Теория и практика судебной экспертизы. 2016. № 1 (41). С. 78–86.

16. Korycki R. Detection of Montage in Lossy Compressed Digital Audio Recordings. *Archives of Acoustics*. 2015. Vol. 39. No. 1. P. 65–72. https://doi.org/10.2478/aoa-2014-0007

16. Korycki R. Detection of Montage in Lossy Compressed Digital Audio Recordings // Archives of Acoustics. 2015. Vol. 39. No. 1. P. 65–72. https://doi.org/10.2478/aoa-2014-0007

ABOUT THE AUTHORS

Boyarov Alexander Grigorievich – Senior State Forensic Expert at Forensic Video and Audio Laboratory of the Federal Centre of Forensic Science of the Russian Ministry of Justice;
email: sass2002@yandex.ru

Siparov Ivan Sergeevich – State Forensic Expert at North-West Regional Centre of Forensic Science of the Russian Ministry of Justice;
email: ivan@siparov.ru

ИНФОРМАЦИЯ ОБ АВТОРАХ

Бояров Александр Григорьевич – старший государственный судебный эксперт лаборатории судебной экспертизы видео- и звукозаписей Российского федерального центра судебной экспертизы при Министерстве юстиции Российской Федерации;
email: sass2002@yandex.ru

Сипаров Иван Сергеевич – государственный судебный эксперт Северо-Западного регионального центра судебной экспертизы Министерства юстиции Российской Федерации;
email: ivan@siparov.ru

Статья поступила: 10.09.2019
Received: 10.09.2019