



METADATA IN AUDIO FILES COMPLIANT WITH ISO/IEC 14496-12 AND THEIR CHARACTERISTICS AS WELL AS THE EVALUATION OF USABILITY IN THE INVESTIGATION OF THE AUTHENTICITY OF RECORDINGS

Marcin MICHAŁEK

Institute of Forensic Research, Kraków, Poland

Abstract

In recent years evidence recordings examined during authentication analysis are recorded in a digital form. Currently a majority of these recordings is being made with mobile devices, mainly mobile phones. These devices usually capture recordings in compressed formats such as 3GP and MP4 in the version with audio only, i.e. M4A. In these types of files data is saved in accordance with the ISO/IEC 14496-12 standard ISO Base Media File Format. Therefore, a significant amount of additional information is stored in the form of so-called metadata besides encoded multimedia in the 3GP and MP4(M4A) file formats. Analysis of the metadata in the file is currently one of the basic methods in process of the authenticity analysis of digital recordings. The structure of files compliant with the mentioned standard consists of objects called boxes containing metadata with information on the recording format and parameters as well as the date and time of the registration. During the work audio recordings in the 3GP and MP4(M4A) formats were recorded using mobile devices, and then the structure of these files was analysed with regard to the ISO/IEC 14496-12 standard. The results of the research show that the metadata of the 3GP and MP4(M4A) file formats contain relevant and useful information for the authenticity analysis of recordings stored in such type of files.

Keywords

Authenticity analysis; Digital recording; File structure; ISO/IEC 14496-12.

Received 28 February 2019; accepted 27 May 2019

1. Introduction

Two global organisations, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), jointly develop and publish standards in electrical and electronic techniques and related sciences (ISO/IEC, 2010; IEC, 2018; ISO, 2018). These organisations established a technical committee named Joint Technical Committee 1 (ISO/IEC JTC 1) in 1987, which aims to establish and develop international standards for Information and Communications Technology (ICT). About 20 smaller subcommittees (SCs) work

within the ISO/IEC JTC 1 committee, including the ISO/IEC JTC 1/SC 29 Coding of audio, picture, multimedia and hypermedia information (ICT, 2014).

ISO/IEC JTC 1/SC 29 subcommittee develops standards and documentation for media encoding and transmission. They have developed the ISO/IEC 14496 standard under the common title Information technology – Coding of audio-visual objects, currently consisting of 33 parts numbered from ISO/IEC 14496-1 to ISO/IEC 14496-33. The standard describes the methods of compression of digital audio and video for recording on data carriers or for transmission in information and communication networks (ISO/IEC, 2010). ISO/IEC 14496 can be divided into several thematic areas, including the

presentation of time-based media and a flexible file format for the exchange, management, editing and presenting of the recorded media (ISO/IEC, 2010).

In accordance with ISO/IEC 14496, the audio or video signal intended for recording or data transmission is first compressed. Additional information (e.g. synchronization) is added to the compressed signal, and then the whole is multiplexed to one or more encoded binary streams, which are either saved to a file or transmitted. When the coded data is read, it is demultiplexed, decompressed and the original signal is reconstructed based on additional information, and then it can be presented to the user (ISO/IEC, 2010).

ISO/IEC 14496 standard is officially related to a very popular MPEG-4 standard and the ISO/IEC Moving Picture Experts Group (MPEG) formally designated ISO/IEC JTC1/SC29/WG11. MPEG-4 was initiated in 1998 and defines compression methods for video and audio data for the whole group of different formats for recording or transferring data. It is used in applications used to compress multimedia in Internet resources when transmitting speech and broadband TV signals, as well as for storing such data in files. It should be added that the file format described in ISO/IEC 14496 is based on the Quick Time format developed by Apple (ISO, 2006). The MPEG-4 standard introduces a number of current technologies and guidelines for both hardware and application manufacturers as well as for users. It enables flexible design of multimedia objects that are adaptable and optimally utilize hardware capabilities to ensure the best possible quality of the data stream (ISO/IEC, 2003; ISO/IEC, 2010; ISO/IEC, 2015).

This paper focuses on part 12 of ISO/IEC 14496, that is, ISO/IEC 14496-12, which describes the format for storing data in ISO-compliant files. Audio recordings recorded with mobile devices of various brands were used as the material for the research, and the section of the standard applicable to audio data has been used for their analysis.

In the age of digitisation and digital signal processing, audio recordings, which are subjected to investigation of authenticity, are recorded in a digital form. The issue of authenticity of digital recordings and current investigative methods have been described quite well in the literature (Brixen, 2007; Grigoras, 2005; Grigoras, Cooper, Michałek, 2009; Hong, Yin, 2013; Koenig, Lacey, 2009; Korycki, 2014; Korycki, 2016; Michałek, 2016; SWGDE, 2018; Zakariah, Khan, Malik, 2018). It should only be mentioned that the investigation of authenticity includes the analysis of both the continuity of the recording and its originality. This is a complex process and global conclusions are drawn

from the results of various phonoscopic investigative methods and, more and more often, with the participation of forensic information technology. Multimedia files contain metadata in their structure that describe valid data with the encoded recording. The analysis of data structure in files is currently one of the basic methods used in the investigation of the authenticity of digital recordings (Kajstura, Michałek, Trawińska, 2017).

2. Organization of data in files compliant with the ISO/IEC 14496-12 standard

ISO/IEC 14496-12 standard promotes a flexible format for the recording of multimedia data in different applications, which will provide for the exchange, editing and presentation of this data to the user. Data can be presented locally by saving to a file (*interchange* option) or remotely via telecommunications networks (*streamed* option). The former ensures free file exchange and archiving, while the latter – simultaneous reception by any number of users. This standard provides guidelines for programmers encoding multimedia data according to this format, called ISO Base Media File Format (ISO/IEC, 2015).

ISO/IEC 14496-12 standard describes the basic file architecture and design recommendations that underlie other standards and storage formats. The file format given in the standard contains rules for time synchronisation, file structure and additional information for sequentially saved multimedia data such as video and audio presentations (ISO/IEC, 2015). The structure of the 3GP and MP4(M4A) format files which have been registered as the research material is based on the guidelines of this standard ISO/IEC (3GPP, 2003; 3GPP2, 2007; ISO/IEC, 2003; ISO/IEC, 2015).

A file saved according to the rules set out in ISO/IEC 14496-12 is a so-called data container for many applications. It can store various types of media, such as video and audio tracks, auxiliary data, text tracks, video subtitles or fonts. In addition, each one of them can be encoded in a different way. For example, it is possible to save an audio track using the Adaptive Multi-Rate (AMR) or Advanced Audio Coding (AAC) encoding in a 3GP file.

According to the ISO/IEC 14496-12 standard, the structure of each file consists of a number of objects that are called boxes. The term “atom” is used in some technical specifications instead of the “box” term, e.g. in the first definition of the MP4 format (ISO/IEC, 2015). Each box is a separate part of data, and there are obligatory and optional boxes in each file. All data in

the file, that is, both multimedia and auxiliary data, are contained in the boxes, and the standard does not provide for the existence of another form of data storage. Each box starts with a header with information about its size and the type of the box. What is important, data in boxes is stored in the big-endian convention, i.e. the most significant byte as the first one, in contrast to the popular RIFF media storage format, where the reverse order, i.e. little-endian, applies. Each box can contain other boxes inside, which creates an ordered and hierarchical file structure. ISO/IEC 14496-12 standard distinguishes between approximately 100 structured boxes containing information fields (metadata) or multimedia data.

The box size is an integer that indicates how many bytes there are in the box. This value specifies its total content, i.e. both the header and data fields and any other boxes inside the main box. The box size can be presented in a compact or extended form, i.e. written using 32 or 64 bits. Currently, the typical box size is written using 32 bits, which is sufficient. Only boxes containing a lot of multimedia data or encoded date and time may require 64 bits.

The box type identifies the object in the file and specifies the type of information stored in it. It indicates whether the box contains encoded multimedia data or metadata, and if so, what kind of data it is. The box type is usually written in a compact format as 4 alphanumeric characters for easier identification in ASCII code. Many boxes also contain additional

information such as the version number and the so-called flags directly after the indication of their type. The box version number, usually 0 or 1, usually indicates whether the box size is compact or extended. Flags, on the other hand, indicate the enablement of additional options or information within the box. An example is a notification that the track is available to the user, it is in the current file and can be played back. Figure 1 shows the general structure of a single box in a file according to ISO/IEC 14496-12 valid for all boxes in each file saved according to this standard.

A file format compliant with this standard shall also include the case where the size of the multimedia data is significant and exceeds the extent of the single file. Then one file contains only metadata, and the subsequent files contain multimedia for presentation.

A typical multimedia file saved according to a standard is self-contained and contains all the necessary information to reproduce the multimedia it contains. It contains the main boxes that define the file type (File Type Box), encoded data (Media Data Box), and metadata (Movie Box). Figure 2 shows the basic structure of the ISO/IEC 14496-12 compliant file.

The file structure may also contain other main boxes, including those defining fragmented multimedia data (Movie Fragment(s) Box) or free space in the file (Free Space Box). In order to ensure greater efficiency in the process of reading information, ISO/IEC 14496-12 defines the order of main boxes in the file structure. The most relevant are listed below.

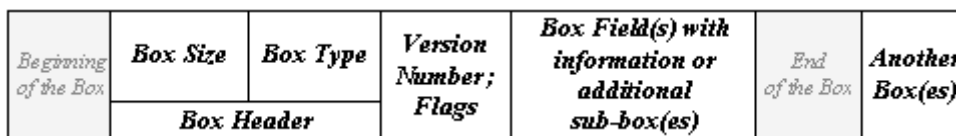


Fig. 1. General structure of a single file box compliant with the ISO/IEC 14496-12 standard. The box size is the entire size of the box with the header, fields, and all contained sub-boxes.

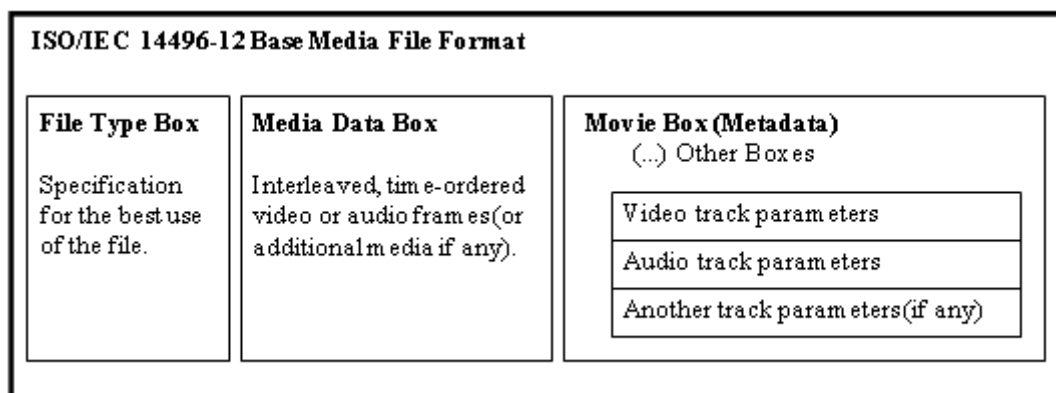


Fig. 2. An example of a simple interchange multimedia file compliant with the ISO/IEC 14496-12 standard.

It is recommended that the box specifying the file type (File Type Box) be placed at the very beginning, before the other main boxes. In turn, the box with multimedia metadata (Movie Box) should be located at the end of the file. In larger main boxes there are additional smaller boxes with headers (Header Boxes), which must be placed first within the main box. If the multimedia data is fragmented, it is recommended that the subsequent fragments in one or more files be arranged according to their order. Figure 3 shows a simplified data structure in an example file compliant with the discussed standard. Simplification consists in the presentation of only the most important boxes in order to depict the idea of such a structure. Some boxes, such as Movie Box, have an extensive hierarchy spanning up to a few levels deep.

Taking into account the possibility of designing own objects, the standard informs that if one of the boxes is not recognized, it should be ignored and the data reading process should be continued from the next recognized box. Some boxes, such as those related to free space (Free Space Box), can even be deleted from a file without affecting the capability to play back saved multimedia.

The metadata contained in the file shall include entries with information related to recording parameters, timestamps or recording equipment applicable for the investigation of recording authenticity (Kajstura et al., 2017; Michalek, 2017). For files stored according to the ISO/IEC 14496-12 standard, the most important object is the box named Movie Box, which contains multimedia metadata. This object will be discussed in detail later in the paper.

The purpose of this paper is to determine what information is contained within audio files recorded in accordance with the ISO/IEC 14496-12 standard, along with an assessment of their applicability in the investigation of the authenticity of recordings. Additionally, we have developed our own tool for automatic analysis of the structure of such files and presentation of their properties.

3. Material and methods for research

In order to carry out the intended research, the recording capabilities of mobile devices of the following brands and models were collected and analysed:

- Mobile phones: LG L9, MANTA MSP 4505, Samsung Galaxy Ace 2 CM11, Samsung Galaxy Core Plus SM-G350, Samsung Galaxy Ace 3 LTE GT-S7275R, Samsung Galaxy Core Prime SM-G361F, Samsung GT-I9060 Galaxy Grand Neo, Samsung GT-C3350, Samsung Galaxy J3, Sony Xperia J1, Sony Ericsson WT19i;
- ASUS Memo Pad HD7 tablet.

105 audio recordings (i.e. without video) were recorded in 3GP and MP4(M4A) formats using these devices. In the further part of the paper, MP4 format files with an audio track will be depicted as M4A files. Pre-installed applications for recording of audio and writing to file were used to produce recordings. These recordings were made using the available functions during recording, i.e. start, stop, and pause the recording. Files with recordings were saved both in the memory of the devices and on the removable microSD

File Type Box			<i>file type and compatibility</i>
Media Data Box			<i>media data container</i>
Movie Box (Metadata)			<i>container for all the metadata</i>
	Movie Header Box		<i>overall declarations</i>
	Track Box		<i>container for an individual track or stream</i>
		Track Header	<i>overall information about the track</i>
		Media Box	<i>container for the media information in a track</i>
	Another Track(s) (if any)		
Free Space Box			<i>free space</i>

Fig. 3. Simplified structure of an exemplary file stored in accordance with the ISO/IEC 14496-12 standard.

memory card. To safely explore the content of memory with recordings, a device was used that prevents the modification of data on this type of media, i.e. the write-blocker. For further analysis, temporary copies of files from source media to a hard disk of the laboratory computer were made.

In order to assess the possibility of using the information stored in the structure of recorded files to investigate the authenticity of recordings, the following analyses were carried out:

- Properties of files read using the Windows 7 Ultimate operating system and two programs designed to analyse multimedia files, i.e. `ffprobe v. N-90893-gcae6f806a6` and `MediaInfo v. 17.10`;
- Visualization, reading and interpretation of data structure in recorded files in order to identify individual objects (boxes) based on the ISO/IEC 14496-12 standard and using hexadecimal editors `010 Editor v. 6.0.2` and `HxD v. 1.7.7.0`;
- Determination of the content of information (metadata) in the disclosed boxes and its significance for the analysis of authenticity.

4. Results

4.1. Analysis of file properties

The purpose of this part of the investigation was to determine the possibility of reading and interpreting the parameters of 3GP and M4A audio files using standard tools. At the beginning, the analysis of the information presented by the Windows 7 Ultimate operating system was performed. It was found that this system shows basic information about both types of files: name with extension, size, timestamps, as well as approximate values of bit rate and recording duration. For files in 3GP format, it also shows the number of channels and the approximate value of the audio sampling rate, and classifies them as video files as well.

Dedicated programs such as `ffprobe` and `MediaInfo` are also used to read information about multimedia files. These are frequently used tools that provide more precise information about such files. They recognize many media recording formats from popular to those rarely used. In addition to basic information available in Windows, these two programs also present detailed parameters of recorded audio in 3GP and M4A files: format, format profile, encoding, sampling rate, bitrate, duration and timestamps generally for the file and audio data. Based on the summary of information read using these tools, it has been established that the audio files recorded are characterised by the following significant parameters:

- 3GP files have `*.3gpp` extensions, while M4A files have `*.m4a` extensions;
- Sound has been recorded in a compressed form in all files: AAC-LC or AMR-NB coding was used in 3GP files, while AAC-LC coding was used in M4A files;
- Bitrates were in the range of 32-156 kbps (AAC-LC) and 12 kbps (AMR-NB) for 3GP files, and in the range of 56-128 kbps for M4A files;
- Sampling rates were 44100 Hz (AAC-LC) or 8000 Hz (AMR-NB) for 3GP files, and 16000-48000 Hz for M4A files;
- Depending on the selected options of `ffprobe` and `MediaInfo` programs, the duration of recordings is presented with different precision; what is important, timestamps are read by these programs from the metadata in the file (creation and modification in `MediaInfo`, creation only in `ffprobe`).

4.2. Analysis of information contained in the metadata

ISO/IEC 14496-12 standard describes the architecture of multimedia files and defines the meaning and structure of objects contained in them called boxes. This paper reviews and describes the standard-compliant boxes included in recorded 3GP and M4A files containing important metadata for the investigation of the authenticity of such recordings. In addition, the figures show examples in the form of visualisations of boxes in hexadecimal and ASCII codes. Each significant box and its fields have been described in detail to be used in a constructed computer program for automatic analysis of files based on the ISO/IEC 14496-12 standard.

File Type Box

Multimedia files compliant with the ISO/IEC 14496-12 standard may be compatible with one or more detailed technical specifications. The File Type Box is one of the main and mandatory objects in a file and should be located at the very beginning of its structure, in front of the other main boxes. It specifies the file type and provides additional information about supported and compatible specifications. The information in this box may indicate the type of encoding used, the file extension, requirements for the support of specific objects, or its general purpose. The discussed box contains information relevant for the playback applications in order to obtain the best possible quality of data readout, as well as the information

which is important in the process of inspecting the file and removing any errors.

The File Type object in the Box Type field of the header contains the *ftyp* character string, which identifies this box in the structure of the entire file. This box contains the following fields:

- *major_brand*: Identifies the best specification for file handling;
- *minor_version*: Version of this specification;
- *compatible_brands*: Compatible specifications.

The ISO/IEC 14496-12 standard provides examples of a dozen or so specifications that can be implemented in the File Type object, including: ISO Base Media File version 1 (labelled *isom*) and in higher versions (for example, version 2 as *iso2*), AVC (labelled *avc1*) and MPEG v. 7 (*mp71*). It is of course possible to implement other specifications provided that they are defined. Higher versions of specification data introduce support for additional functions or boxes.

The File Type Box was identified in the structure of all collected 3GP and M4A files at the beginning of the data.

For 3GP files, the *major_brand* fields pointed to the specifications of 3GPP v. 4 (*3gp4* entry), while *compatible_brands* pointed to ISO Base Media File v. 1 and 3GPP v. 4.

For M4A files, the *major_brand* fields pointed to the specifications of MP4 v. 2 (*mp42* entry) and 3GPP v. 4, and *compatible_brands* pointed to ISO Base Media File v. 1, 3GPP v. 4, and MP4 v. 2.

In case of the vast majority of 3GP and M4A files the size of the whole File Type Box was 24 bytes and included one *major_brand* entry and two *compatible_brands* entries. In turn, M4A files recorded with Samsung GT-C3350 were 28 bytes in size and three *compatible_brands* entries, which is a feature that distinguishes this phone model from the others and additional information for the investigation of authenticity. None of the analysed File Type Boxes had internal boxes in its structure. Figure 4 shows a fragment of the data structure in the 3GP format file with the box File Type selected.

Files compliant with older specifications may not contain the File Type Box. The current standard provides for such a case and indicates that when reading such files, it should take the values *major_brand* and *compatible_brands* as *mp41*, and the value *minor_version* as 0.

Encoded multimedia, Media Data Box

Encoded multimedia data is contained in the Media Data Box. This is one of the main boxes in the file structure, although it is not mandatory. According to the standard, there can be a file containing only metadata, while the associated media are contained in separate files. Absolute addressing, i.e. counted as an offset from the beginning of the data in the entire file, is mandatory for metadata describing multimedia data. It is also possible that one file will contain several Media Data Boxes.

Analysis of the data structure of recorded 3GP and M4A files showed that each of them contained exactly one box with multimedia. Media Data Box contains a string of characters *mdat* in the 8-byte header that identifies it within the file. The header also contains information about the size of the box, which depends on the size of the multimedia data. This box had no other internal boxes in its structure in the recorded audio files. It was the most extensive object in the whole multimedia file for the analysed recordings.

The structure of the Media Data Box is not complicated compared to other objects, but it allows for quite a significant analysis. The size of the header and the declared size of the entire *mdat* box are known. It is also possible to read the real size of the multimedia data itself directly from the hexadecimal record and compare it to the value given in the header. The mismatch between these two values could, for example, result from the deletion of some data. According to the ISO/IEC file structure described above, it is also possible to investigate whether the header of another box is located directly after the Media Data Box, which would indicate a correct structure. This provides addi-

Offset (h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	00	00	00	18	66	74	79	70	33	67	70	34	00	00	00	00ftyp3gp4....
00000010	69	73	6F	6D	33	67	70	34	00	00	0C	00	66	72	65	65	isom3gp4....free
00000020	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000030	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000040	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00

Fig. 4. A section of data beginning in the 3GP format file with highlighted File Type Box. Visible poles are: 3GP v. 4 as a *major_brand* and ISO Base Media File v. 1 or 3GP v. 4 as *compatible_brands*.

tional verification of the size of the declared multimedia data. Based on the hexadecimal analysis of the collected 3GP and M4A files, it was established that the header of the box with metadata was located directly behind the multimedia data. Fig. 5 shows a fragment of the M4A file structure with the header of the *mdat* box selected.

Metadata in the file, Movie Box

The box named Movie Box is the most important object for the investigation of the authenticity of recordings stored in files according to this standard. This is one of the main and mandatory objects in the file structure that contains metadata for the entire data presentation. As recommended by the standard, the box with metadata should be located at the end of the entire file.

Based on the analysis of the structure of the collected 3GP and M4A files, it was found that Movie Box is a very complex object with a hierarchical and multi-level structure containing numerous internal boxes. It starts with a typical 8-byte box header with a string of characters *moov* as an identifier. There is also a field in this header with the size of the entire box, the value of which is not constant. It depends on many factors, especially the number of boxes determined by the designer, the detail of descriptions, the number of tracks, and even the size of the multimedia data. It is usually one of the most extensive objects in relation to the other main boxes. It contains from

several dozen to several hundred kilobytes on average, and even more than a megabyte of information in the case of longer recordings (which has been determined on the basis of recordings with typical AAC-LC audio encoding, 44100 Hz sampling rate, and lasting from several dozen seconds to 2 hours). Exactly one Movie Box object was found in each 3GP and M4A file.

There are internal boxes, both mandatory and optional, within the Movie Box. The meaning and content of the most important ones will be discussed below.

Header of the box with metadata, Movie Header Box

Due to the size and complexity of the Movie Box object, it contains an additional header in the form of an internal Movie Header Box. A single object of this type was found in the structure of each analysed 3GP and M4A file. It is located directly behind the 8-byte header at the beginning of the main box, i.e. Movie Box, and is its mandatory element. Movie Header contains an entry of *mvhd*, which is its identifier. This box contains general information about the entire presentation of multimedia data. Based on the standard and collected 3GP and M4A files, it has been determined that it contains the following relevant fields:

- *version*: Box version, it may have a value of 0 or 1; if the value is 0, the following fields described below *creation_time*, *modification_time*, and *duration*

Offset (h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	00	00	00	18	66	74	79	70	33	67	70	34	00	00	00	00ftyp3gp4....
00000010	69	73	6F	6D	33	67	70	34	00	1E	DC	40	6D	64	61	74	isom3gp4.Ü@mdat
00000020	00	00	18	83	F2	1C	00	00	18	83	F2	1C	00	00	18	83ň.....ň.....
00000030	F2	1C	00	00	18	83	F2	1C	00	00	18	83	F2	1C	00	00	ň.....ň.....ň...
00000040	18	83	F2	1C	00	00	18	83	F2	1C	00	00	18	83	F2	1C	..ň.....ň.....ň.
00000050	00	00	18	83	F2	1C	00	00	18	83	F2	1C	00	00	18	83ň.....ň.....

a)

001EDC00	D0	30	B6	4A	E5	0E	21	66	2A	18	5A	A0	28	84	EC	22	Đ09JÍ. !f*.Z (,ě"
001EDC10	CF	89	07	00	02	EA	20	0E	44	34	0F	1C	32	49	FC	C7	Đ%...ē .D4..2IüÇ
001EDC20	17	52	AF	D9	23	F8	E6	1D	55	E5	F4	D4	B1	0D	3A	39	.RZÜ#řé.UíóÔ±.:9
001EDC30	DA	D7	B0	22	9A	0A	00	9C	8C	62	28	83	04	20	4E	18	Ú×°"š.ĐésŠb(..N.
001EDC40	40	00	C3	19	06	31	00	D9	90	1B	3F	9F	98	15	B7	60	ž.Ă..1.Ů..?ž...'
001EDC50	34	F8	7B	01	7A	30	00	0E	00	00	5A	6F	6D	6F	6F	76	4ř(.z0...Zomoov
001EDC60	00	00	00	6C	6D	76	68	64	00	00	00	00	D5	7B	BB	2D	...lmvhd....Ů{»-

b)

Fig. 5. The header of the Media Data Box: 00 1C CA D5 bytes in hexadecimal system is the size of the whole box (1,886,933 bytes in decimal system) and mdat is the identifier of this box (Figure 5, part a). The ending fragment of the Media Data Box; directly after those data the header of the Movie Box is situated (Figure 5, part b).

- have a size of 32 bits, and if the value is 1, they have a size of 64 bits;
- *creation_time* and *modification_time*: The value associated, respectively, with the creation and modification of multimedia presentations; it is counted every 1 second from midnight on 1 January 1904 and presented as Universal Time Coordinated (UTC);
- *timescale*: The number of time units in 1 second;
- *duration*: The duration of the presentation in seconds for a defined time scale equal to the duration of the longest track in the entire presentation;
- *next_track_ID*: The identifier of the next track in the presentation.

The analysis of all recorded files with recordings allowed us to additionally determine that the values of the following fields were equal: *version* – 0, *timescale* – 1000 (i.e. time counted every 1 millisecond), *next_track_ID* – 2, and the parameter *duration* was equal to the duration of the entire recording.

Particularly noteworthy are the following fields: *creation_time* and *modification_time*. They contain the number of seconds in the hexadecimal representation to be decoded into the decimal system, formatted into date and time and calculated as a difference from midnight on 1 January 1904. This enables to determine the date and time of creation and modification. Based on the study it was established that the calculated pairs of creation and modification times for each 3GP and M4A file were the same and related to the time the recording ended. The investigation of authenticity should also take into account the difference between Central European Time and UTC, which is +1 or +2 hours if daylight saving time applies. Fig. 6 shows the data structure in the M4A file with the selected Movie Header fragment containing the entries from which it is possible to determine the discussed timestamps for the presentation.

For the analysed files, it was determined that the calculated timestamps depend on the system clock set-

ting in the device, and the menu of the analysed devices allowed the user to change this parameter. However, mobile devices, such as mobile phones and tablets and smartwatches with phone functionality, have the default option to automatically update the time that is provided by the network operator. If they are within the network range, then the set time should be up-to-date, including the one encoded in the Movie Header Box. Where possible, timestamps determined on the basis of the metadata mentioned above should be verified by other means, for example using the powerline hum signal.

The times calculated based on the entries in the *creation_time* and *modification_time* fields are independent of the Windows operating system indications and do not change when copying the file to another storage medium, which is crucial for the authenticity analysis. Often only copies of recordings on removable media, which are characterized by modified file system timestamps, are transferred for investigation.

Tracks in the multimedia file, Track Box

Multimedia files can contain many tracks, such as video or audio. Their description is located in the Track Box objects, and a separate box of this type is used for each single track. This is one of the basic and mandatory objects in the main Movie Box. Exactly one audio track was found in each of the analysed 3GP and M4A files, recorded in one Track Box object. The identifier of this box consists of a string of characters *trak* located in its header.

The Track Box object has an additional header Track Header Box as an internal, single and mandatory object with the *tkhd* identifier. This header contains general metadata for the saved track and includes the following relevant fields:

- *version*: The value of 0 or 1; indicates the size of the timestamp and the duration of the track on 32 or 64 bits, respectively;

Offset (h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00086910	03	D9	FA	AD	1C	F8	F0	07	00	00	1A	63	6D	6F	6F	76	.Ùú.řđ...cmoov
00086920	00	00	00	6C	6D	76	68	64	00	00	00	00	D5	7B	BC	EA	...lmvhd...Ń{Le
00086930	D5	7B	BC	EA	00	00	03	E8	00	00	8B	0C	00	01	00	00	Ń{Le...č...<.....
00086940	01	00	00	00	00	00	00	00	00	00	00	00	00	01	00	00
00086950	00	00	00	00	00	00	00	00	00	00	00	00	00	01	00	00
00086960	00	00	00	00	00	00	00	00	00	00	00	00	40	00	00	00@...
00086970	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00

Fig. 6. A section of the Movie Header Box: two consecutive hexadecimal values D5 7B BC EA allow to calculate the date and time of creation and modification, respectively. After decoding the date and time is equal to 2017-06-30 08:31:38 UTC [yyyy-mm-dd hh:mm:ss].

- *creation_time* and *modification_time*: The value associated with creating and modifying the track; as in Movie Header, the time is counted every 1 second from midnight on 1 January 1904 and shown as Universal Time Coordinated(UTC);
- *track_ID*: A unique value that identifies the track in the file;
- *duration*: The track duration for the defined timescale in Movie Header;
- *flags*: The field containing so-called flags that define the state of the track, that is, whether it is available or locked to the user, determine if the track is used in the presentation and can be played back.

Based on the collected 3GP and M4A recordings, it was determined that the timestamps from the creation and modification of the track and its duration were consistent with the corresponding values for the entire presentation. The track duration for each analysed file was expressed in milliseconds. The values of the following fields were: *version* – 0, *track_ID* – 1 (i.e. the first track in the file), but *flags* – 7 (typical value as the sum of consecutive flags), indicating that the track is: available, used in the presentation, and can be played back.

Information about the multimedia in the track, Media Box

Getting deeper into the Movie Box hierarchy of metadata, the box containing information related to the media stored in the track, that is, Media Box, should be distinguished. This is a mandatory and internal box within the Track Box. There was one such object in each 3GP and M4A file that had a string of *mdia* characters in its header.

Like the previously discussed objects, Media Box has its own internal header named Media Header Box and ID *mdhd* with general media information in the track. It was found to contain the following relevant fields:

- *version*: Indicates the size of the timestamps and media duration (0 for 32, 1 for 64 bits);
- *creation_time* and *modification_time*: Timestamps of media creation and modification, whose principle of determination is the same as for the track in the file;
- *timescale*: The value of the time base for recorded media;
- *duration*: The duration of the media in the track in relation to the *timescale* parameter;
- *language*: The language used for the media.

The timestamps of creation and modification as well as the duration of media in the analysed audio

files were in accordance with the respective timestamps and duration for the tracks. The *timescale* parameter, despite its name coinciding with that in Movie Header, was equal in this case to the sampling rate of audio. For these recordings, this parameter ranged from 8000 to 48000. The information in the *language* field is written using three alphanumeric characters, the meaning of which is described in the ISO 639-2/T standard. In the collected 3GP and M4A audio files, the value in this field always pointed to the English language, despite the setting of the Polish menu in the tested devices.

There is a box called Handler Reference Box inside the Media Box object, which points to the media track type and thus the way it is played back. It is identified by the string *hdlr* and has the following relevant fields:

- *handler_type*: Indicates the type of track with saved media;
- *name*: Verbal description of the track type.

The 3GP and M4A files recorded for the investigation contained audio tracks and were marked with entries: the field *handler_type* – *sound*, while the field *name* – *SoundHandle*.

There is also the Media Information Box object within Media Box identified by the character string *minf*, which contains additional internal boxes with information important for analysing the authenticity of recordings. There are boxes marked as Sample Description Table and Bit Rate Box in the Sample Table Box object with the *stbl* identifier. The structure of the first one depends on the track type (i.e. the value of *handler_type* described above) and contains the *format* field with information about the media saving format. For the collected audio recordings, this box was identified by a string of characters *stsd*, and the format field included entries: *mp4a* (MP4 for audio, AAC-LC encoding profile) or *samr* (Adaptive Multi-Rate in Narrowband version) for 3GP files, and exclusively *mp4a* for M4A files. As one can see, the audio in the analysed recordings was saved with lossy compression only. The second object is Bit Rate Box identified by the *btrt* character string. It contains fields *maxBitrate* and *avgBitrate*, which represent the maximum and average bit rate values in bits per second for the data stream. The Sample Table Box object also contains information about the time indices of the samples stored in the track and the difference between the current index and the beginning of the track. The Decoding Time to Sample Box with identifier *stts* contains a compact table with time indices assigned to sample numbers. According to the standard, a sample means a single section of encoded data (a frame) that can be assigned a time index. The analysis of the collected

3GP and M4A recordings showed that the box usually contains two pairs of entries: the first one with the number and index for the first sample, the second one with the values for the last sample. One can determine the duration of all samples by the number of samples and their indices over time. The calculated value should match the duration of the media and allow for additional verification with the parameters stored in the file structure. This match has been established for the recorded 3GP and M4A files. Fig. 7 shows an example of parameters stored in the *stts* box.

Free space, Free Space Box

ISO/IEC 14496-12 standard also defines an object named Free Space Box, which can be one of the main boxes in the file structure hierarchy or an internal box. It does not contain any relevant information and may be omitted during playback of the content of the presentation. It is even possible to delete it from the file without affecting the saved media. Its structure is uncomplicated and contains a header with the *free* identifier and data. However, the analysis of the collected files has allowed the observation that this optional box is not very common. It was seen only in 3GP files recorded with Asus, LG, Manta, and Samsung Galaxy Ace 2 devices. It was located between the box defining the file type and the box with multimedia, and its data was a sequence of bytes filled with zeroes. Despite its insignificant content, Free Space Box is a certain distinguishing feature of these devices, which may be additional information in the process of investigation of authenticity.

Table 1 summarizes the key metadata that was read from the structure of the analysed 3GP and M4A audio files, taking into account the research results described above.

Table 1
Comparison of the key metadata in analysed 3GP and M4A audio files

box: field	metadata in files	
	3GP	M4A
ftyp: major_brand	3gp4	mp42 or 3gp4
ftyp: compatible_brands	isom or 3gp4	isom, 3gp4 or mp42
mdat	amount of the multimedia	
mvhd: version	0	
mvhd: creation_time	32 bits; depends on the time settings	
mvhd: modification_time		
mvhd: duration	32 bits; depends on the length of the presentation	
tkhd: creation_time	32 bits; depends on the time settings	
tkhd: modification_time		
tkhd: track_ID	1	
tkhd: duration	32 bits; depends on the length of the track	
mdhd: creation_time	32 bits; depends on the time settings	
mdhd: modification_time		
mdhd: timescale	8000 or 44100	16000–48000
mdhd: duration	32 bits; depends on the media length in the track	
mdhd: language	15 bits; coded using the ISO 639-2/T	
hdlr: handler_type	soun	
hdlr: name	SoundHandle	
stsd: format (coding type)	mp4a or amr	mp4a
stts: sample_count	32 bits; usually 2 entries	
stts: sample_delta (in timescale)	32 bits; usually 2 entries	
free	occur; a header and 0's	no entries

4.3. Software for the automatic analysis of file and metadata structure

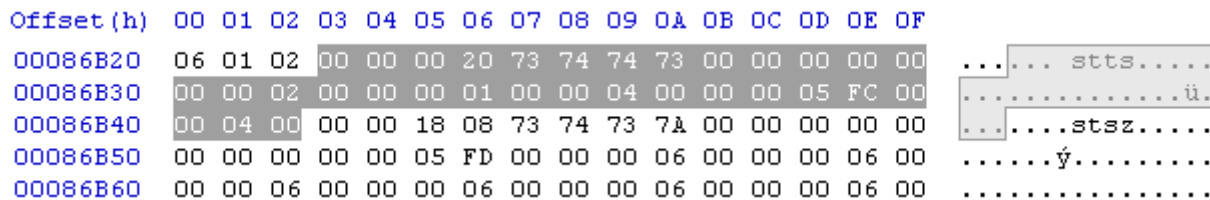


Fig. 7. The Decoding Time to Sample Box: by multiplying the number of consecutive samples (00 00 05 FC in hexadecimal) and the value of intervals between them (00 00 04 00) with respect to the media timescale (00 00 AC 44) it is possible to calculate the duration of all samples (35 seconds).

Algorithms for automatic analysis of this type of files were developed based on the obtained results of the study of the 3GP and M4A file structure as well as the ISO/IEC 14496-12 standards. A detailed interpretation of the fields and metadata of all the boxes described in the paper has been made and a database has been created with a description of the meaning of each box for the purposes of the created software. A computer program has been created using MATLAB computing environment, whose general principle of operation is as follows:

1) Import the whole file to the program workspace;

- 2) Determine the type of recorded tracks (video, audio or auxiliary);
- 3) Verify of the declared against calculated size of multimedia data;
- 4) Analyse the structure of the imported file and detect subsequent boxes based on their identifiers;
- 5) Read hexadecimal and ASCII data and metadata from boxes;
- 6) Determine the values of fields and parameters using the built-in database;
- 7) Verify the integrity of the file structure and detect possible disturbances or lack of significant boxes;

```

Command Window

File name: Glos 006.m4a

== 'FTYP' box size: 00000018 [HEX]
'FTYP' major brand | version: 3gp4 [ASCII] | 00000000 [HEX]
'FTYP' compatible brands: isom3gp4 [ASCII]

== 'MDAT' box size: 0010124F [HEX]
Next 8[B] after MDAT data are: 00002FEB 6D6F6F76 [HEX]; " /ëmoov" [ASCII]
The data in the MDAT box is: 'consistent'

== 'MOOV' box size: 00002FEB [HEX]
='MOOV\MVHD' box size: 0000006C [HEX]
'MOOV\MVHD' version [DEC]: '      [0]
'MOOV\MVHD' Presentation: creation time [HEX]:      D57973CF
'MOOV\MVHD' Presentation: modification time [HEX]: D57973CF
'MOOV\MVHD' Presentation: creation time UTC [DATE]: 2017-06-28 14:55:11
'MOOV\MVHD' Presentation: modification time UTC [DATE]: 2017-06-28 14:55:11
'MOOV\MVHD' Presentation: timescale [DEC]: '      [1000]
'MOOV\MVHD' Presentation: duration [s]: '      [67.5930]

='MOOV TRACK(s)
'MOOV\TRAK' There are: '      [1]      ' [TRACKS] in the file.'
'MOOV\TRAK' There are: '      [1]      ' [AUDIO TRACKS].
'MOOV\TRAK' There are: '      [0]      ' [VIDEO TRACKS].
'MOOV\TRAK' There are: '      [0]      ' [AUX VIDEO TRACKS].
'MOOV\TRAK' There are: '      [0]      ' [TEXT TRACKS].
'MOOV\TRAK' There are: '      [0]      ' [SUBTITLES TRACKS].
'MOOV\TRAK' There are: '      [0]      ' [FONT MEDIA TRACKS].
-'MOOV\TRAK' 1st AUD TRACK: size 00002F37 [HEX]
'MOOV\TRAK' 1st AUD TRACK: version '      [0]
'MOOV\TRAK' 1st AUD TRACK: flags (typical=?) '      [7]
'MOOV\TRAK' 1st AUD TRACK: creation time [HEX]:      D57973CF
'MOOV\TRAK' 1st AUD TRACK: modification time [HEX]: D57973CF
'MOOV\TRAK' 1st AUD TRACK: creation time UTC [DATE]: 2017-06-28 14:55:11
'MOOV\TRAK' 1st AUD TRACK: modification time UTC [DATE]: 2017-06-28 14:55:11
'MOOV\TRAK' 1st AUD TRACK: duration (t.scal. MVHD) [s]: '      [67.5930]

```

Fig. 8. Automatic analysis of the metadata in the M4A audio file using the created software – a part of the report containing information on File Type Box, Media Data Box, Movie Header Box and Track Box as well.

8) Generate a report from the file analysis containing the description of boxes and their metadata as well as the extracted parameters.

Fig. 8 shows the fragment of a sample report generated from the metadata analysis in an M4A file using the developed software.

Fig. 9 shows the fragment of the report based on the analysis of metadata in the M4A file where one of the main and mandatory objects, i.e. Movie Box, is damaged.

The development of our own software was motivated by the need to create a tool dedicated to the analysis of authenticity of recordings, which will provide a reliable analysis of the structure of files and boxes with precise naming of parameters in accordance with the principles of the ISO/IEC standard. Due to the flexibility of the structure of this type of files, designers of devices and applications can put new boxes containing multimedia or metadata in the structure. Along with the analysis of subsequent evidence and test recordings, it is possible to easily update the database with new recognized boxes for use in the investigation of authenticity.

5. Conclusions

The research described in this paper enabled to obtain and systematize knowledge about the structure of multimedia files recorded in accordance with the ISO/IEC 14496-12 standard. The studies were carried out due to the increasing number of evidence recordings subjected to authenticity analysis, which are recorded in formats based on the mentioned standard, such as

3GP and MP4(M4A). Based on the collected recordings and the ISO/IEC standard, it was established that the structure of the files of this kind consists of boxes, within which the media and metadata are stored. The analysis and evaluation of these elements has showed that they contain relevant information which can be used to investigate the authenticity of recordings. They enable to determine the validity of the file structure, media recording parameters and timestamps of the recordings. This is important because file structure analysis is currently one of the basic methods used to investigate the authenticity of recordings. Based on the results obtained, the software was developed for the automatic analysis of file structure according to ISO/IEC 14496-12, intended for use in the MATLAB computing environment. It performs a data integrity analysis and determines valid parameters and possible disturbances in the file structure. It is a tool supporting the investigation of authenticity of recordings performed in connection with the opinion-forming activity of the Institute.

References

1. 3rd Generation Partnership Project (3GPP). (2003). Technical Specification 3GPP TS 26.244: Transparent end-to-end packet switched streaming service (PSS); 3GPP file format (3GP) (Release 6). Retrieved January, 23, 2018 from <http://www.3gpp.org>.
2. 3rd Generation Partnership Project 2 (3GPP2). (2007). 3GPP2 C.S0050-B Version 1.0: 3GPP2 File Formats for Multimedia Services. Retrieved October, 17, 2016 from <http://www.3gpp.org>.

```

Command Window

File name: Glos 008.m4a

== 'FTYP' box size: 00000018 [HEX]
'FTYP' major brand | version: 3gp4 [ASCII] | 00000000 [HEX]
'FTYP' compatible brands: isom3gp4 [ASCII]

== 'MDAT' box size: 002C2A12 [HEX]

== 'STCO' box size: 00000014 [HEX]
Name of the box before 'stco' 1-st offset: 6D646174 [HEX]; mdat [ASCII]
All 'stco' offsets are: 00000020 [HEX]

Error using ISO metadata analyzer (line 290)
The header of the 'MOOV' box is damaged or missing!

```

Fig. 9. Detection of a damaged Movie Box in the M4A audio file structure.

3. Brixen, E. (2007). Techniques for the authentication of digital audio recordings. *Audio Engineering Society Convention Paper*, 7014. Retrieved January, 29, 2019 from <https://www.researchgate.net>.
4. International Organization for Standardization (ISO). (2018). *Economic benefits of standards*. Retrieved December, 28, 2018 from <http://www.iso.org>.
5. International Organization for Standardization (ISO). (2006). ISO/IEC JTC 1/SC 29/WG 11 N8150. ISO Base Media File Format white paper. Retrieved January, 29, 2019 from <http://www.chiariglione.org/mpeg/>.
6. International Electrotechnical Commission (IEC). (2018). Welcome to the IEC. Retrieved December, 28, 2018 from <http://www.iec.ch>.
7. Information and Communication Technology (ICT). (2014). ISO/IEC JTC 1: Vision, Mission and Principles. Retrieved December, 28, 2018 from: <http://www.iso.org>.
8. ISO/IEC 14496-14:2003. Information technology – coding of audio-visual objects – Part 14: MP4 File Format.
9. ISO/IEC 14496-1:2010. Information technology – coding of audio-visual objects – Part 1: Systems.
10. ISO/IEC 14496-12:2015. Information technology – coding of audio-visual objects – Part 12: ISO base media file format.
11. Michałek, M. (2017). Properties of recordings and audio files saved in AMR format and an assessment of the possibility of applying them in authenticity examinations. *Problems of Forensic Sciences*, 109, 27–42.
12. Grigoras, C. (2005). Digital audio recording analysis: The electric network frequency (ENF) criterion. *International Journal of Speech, Language and the Law*, 12, 64–76.
13. Grigoras, C., Cooper, A., Michałek, M. (2009). Forensic Speech and Audio Analysis Working Group – Best practice guidelines for ENF analysis in forensic authentication of digital evidence; REF. CODE: FSAAWG-BPM-ENF-001. Retrieved October, 13, 2017 from <http://www.enfsi.eu>.
14. Hong, G., Yin, Z. (2013). Research on digital audio authenticity analysis. Retrieved May, 25, 2016 from <https://www.researchgate.net>
15. Koenig, B., Lacey, D. (2009). Forensic authentication of digital audio recordings. *Journal of Audio Engineering Society*, 57, 662–695.
16. Korycki, R. (2014). Badanie autentyczności cyfrowych nagrań fonicznych utrwalonych w plikach MP3. *Problemy Kryminalistyki*, 283(1), 2–17.
17. Korycki, R. (2016). Problematyka badania autentyczności cyfrowych nagrań fonicznych. *Prokuratura i Prawo*, 12, 138–157.
18. Michałek, M. (2009). The application of powerline hum in digital recording authenticity analysis. *Problems of Forensic Sciences*, 80, 355–364.
19. Michałek, M. (2016). Test audio recordings and their use in authenticity examinations. Database of properties of digital audio recorders and recordings. *Problems of Forensic Sciences*, 105, 355–369.
20. Scientific Working Group on Digital Evidence, SWGDE. (2018). Best practices for digital audio authentication. Retrieved November, 30, 2018 from <https://www.swgde.org>.
21. Zakariah, M., Khan, M., Malik, H. (2018). Digital multimedia audio forensics: past, present and future. *Multimedia Tools and Applications*, 77, 1009–1040.

Corresponding author

Dr Marcin Michałek
Institute of Forensic Research
ul. Westerplatte 9
PL 31-033 Kraków
e-mail: mmichalek@ies.edu.pl

METADANE W PLIKACH DŹWIĘKOWYCH ZGODNYCH Z NORMĄ ISO/IEC 14496-12 ORAZ ICH CHARAKTERYSTYKA I OCENA MOŻLIWOŚCI WYKORZYSTANIA W BADANIACH AUTENTYCZNOŚCI NAGRAŃ

1. Wstęp

Dwie organizacje o zasięgu globalnym, tj. Międzynarodowa Organizacja Normalizacyjna (*International Organization for Standardization*, ISO) oraz Międzynarodowa Komisja Elektrotechniczna (*International Electrotechnical Commission*, IEC) wspólnie opracowują i publikują normy z zakresu technik elektrycznych i elektronicznych oraz nauk powiązanych (ISO/IEC, 2010; IEC, 2018; ISO, 2018). Organizacje te w 1987 roku powołały komitet techniczny o nazwie ISO/IEC *Joint Technical Committee 1* (ISO/IEC JTC 1), którego celem jest opracowywanie i rozwijanie międzynarodowych norm z zakresu technologii informacyjnych i komunikacji (*Information and Communications Technology*, ICT). W obrębie komitetu ISO/IEC JTC 1 pracuje około 20 mniejszych podkomitetów (*subcommittees*, SCs), wśród których działa podkomitet oznaczony ISO/IEC JTC 1/SC 29 *Coding of audio, picture, multimedia and hypermedia information* (ICT, 2014).

Podkomitet ISO/IEC JTC 1/SC 29 zajmuje się tworzeniem standardów oraz dokumentacji związanej z kodowaniem i przesyłaniem multimediów. W ramach swojej działalności opracował normę ISO/IEC 14496 pod wspólnym tytułem *Information technology – Coding of audio-visual objects* składającą się aktualnie z 33 części oznaczonych od ISO/IEC 14496-1 do ISO/IEC 14496-33. Wymieniona norma opisuje metody kompresji cyfrowego dźwięku i wideo do zapisu na nośnikach danych lub do przesyłu w sieciach teleinformatycznych (ISO/IEC, 2010). Normę ISO/IEC 14496 można podzielić na kilka tematycznych obszarów, w tym związanych z prezentacją sygnałów dźwiękowych i wizualnych z podstawą czasu (*time-based media*) oraz elastycznym formatem pliku umożliwiającym jego wymianę, zarządzanie, edycję i prezentację zapisanych multimediów (ISO/IEC, 2010).

Zgodnie z założeniami normy ISO/IEC 14496 sygnał dźwiękowy lub wideo przeznaczony do zapisu albo przesyłu danych w pierwszej kolejności zostaje skompresowany. Do skompresowanego sygnału dodawane są informacje dodatkowe (np. synchronizacja), a następnie całość multipleksowana zostaje do jednego lub więcej zakodowanych strumieni binarnych, które są zapisywane do pliku albo transmitowane. Podczas odczytu zakodowanych danych są one demultipleksowane, dekompresowane i na podstawie informacji dodatkowych zrekon-

struowany zostaje sygnał pierwotny, który może zostać zaprezentowany użytkownikowi (ISO/IEC, 2010).

Norma ISO/IEC 14496 związana jest oficjalnie z bardzo popularnym standardem MPEG-4 oraz z grupą roboczą ISO/IEC *Moving Picture Experts Group* (MPEG) oznaczoną formalnie ISO/IEC JTC1/SC29/WG11. MPEG-4 został zainicjowany w 1998 roku i definiuje metody kompresji danych wideo i dźwiękowych dla całej grupy różnych formatów przeznaczonych do zapisu lub przesyłu danych. Wykorzystuje się go w aplikacjach służących do kompresji multimediów w zasobach sieci Internet podczas przesyłania sygnału mowy i szerokopasmowej telewizji, jak również do zapisu tego rodzaju danych w plikach. Dodać należy, że format pliku opisany w normie ISO/IEC 14496 oparty został na podstawie formatu *Quick Time* opracowanego przez *Apple* (ISO, 2006). Standard MPEG-4 wprowadza szereg aktualnych technologii i wytycznych zarówno dla producentów sprzętu i aplikacji, jak również dla użytkowników. Umożliwia on elastyczne projektowanie obiektów multimedialnych, które mają możliwość adaptacji i optymalnego wykorzystania możliwości sprzętu, w celu zapewnienia jak najlepszej jakości strumienia danych (ISO/IEC, 2003; ISO/IEC, 2010; ISO/IEC, 2015).

W niniejszej pracy skoncentrowano się na części 12 normy ISO/IEC 14496 oznaczonej ISO/IEC 14496-12, która opisuje format zapisu danych w plikach zgodnych z ISO. Jako materiał do badań posłużyły nagrania dźwiękowe zarejestrowane urządzeniami mobilnymi różnych marek, a do ich analizy wykorzystano normę w części poświęconej danym dźwiękowym.

W dobie digitalizacji i cyfrowego przetwarzania sygnałów nagrania dźwiękowe, które poddawane są badaniom autentyczności, zapisywane są w postaci cyfrowej. Zagadnienie autentyczności nagrań cyfrowych i stosowane aktualnie metody badawcze zostały w literaturze dość dobrze opisane (Brixen, 2007; Grigoras, 2005; Grigoras, Cooper, Michałek, 2009; Hong, Yin, 2013; Koenig, Lacey, 2009; Korycki, 2014; Korycki, 2016; Michałek, 2016; SWGDE, 2018; Zakariah, Khan, Malik, 2018). Należy jedynie nadmienić, że w ramach badań autentyczności analizie poddaje się zarówno ciągłość nagrania, jak i jego oryginalność. Jest to złożony proces, a globalne wnioski wysnute zostają na podstawie wyników uzyskanych z różnych fonoskopijnych metod badawczych oraz, coraz częściej, przy udziale informatyki sądowej. Pliki multimedialne zawierają w swojej strukturze metadane,

które opisują właściwe dane z zakodowanym nagraniem. Analiza struktury danych w plikach jest aktualnie jedną z podstawowych metod wykorzystywanych podczas badań autentyczności nagrań cyfrowych (Kajstura, Michałek, Trawińska, 2017).

2. Organizacja danych w plikach zgodnych z normą ISO/IEC 14496-12

Norma ISO/IEC 14496-12 propaguje elastyczny format zapisu danych multimedialnych dla różnych zastosowań, który zapewni wymianę, edycję i prezentację tych danych użytkownikowi. Prezentacja danych może odbywać się lokalnie poprzez zapisanie do pliku (opcja *interchange*) lub na odległość za pośrednictwem sieci teleinformatycznych (opcja *streamed*). Pierwsza możliwość zapewnia swobodną wymianę plików oraz ich archiwizowanie, druga zaś – jednoczesny odbiór przez dowolną liczbę użytkowników. Omawiana norma zawiera wytyczne dla programistów kodujących dane multimedialne zgodnie z tym formatem, nazwanym *ISO Base Media File Format* (ISO/IEC, 2015).

Norma ISO/IEC 14496-12 opisuje podstawową architekturę plików i zalecenia odnośnie do ich projektowania, co stanowi fundament do tworzenia innych norm i formatów zapisu. Format plików podany w normie zawiera reguły dotyczące synchronizacji czasowej, struktury pliku oraz informacji dodatkowych dla sekwencyjnie zapisanych danych multimedialnych, takich jak prezentacje wizualno-dźwiękowe (ISO/IEC, 2015). Struktura plików formatu 3GP oraz MP4(M4A), które zarejestrowano jako materiał do badań, bazuje na wytycznych omawianej normy ISO/IEC (3GPP, 2003; 3GPP2, 2007; ISO/IEC, 2003; ISO/IEC, 2015).

Plik zapisany zgodnie z zasadami wymienionymi w normie ISO/IEC 14496-12 stanowi tzw. kontener z danymi dla wielu zastosowań. Można w nim zapisać różne rodzaje mediów, takie jak ścieżki wideo i dźwiękowe, dane pomocnicze, ścieżki tekstowe, z napisami dla wideo lub zawierające czcionki. Dodatkowo każdą z nich można zakodować w odrębny sposób. Przykładowo – w pliku 3GP możliwe jest zapisanie ścieżki dźwiękowej z wykorzystaniem kodowania *Adaptive Multi-Rate* (AMR) albo *Advanced Audio Coding* (AAC).

Zgodnie z normą ISO/IEC 14496-12 struktura każdego pliku składa się z pewnej liczby obiektów, które nazywane są boksami. Zamiast terminu „boks” w niektórych specyfikacjach technicznych stosowane jest określenie „atom”, na przykład w pierwszej definicji formatu MP4 (ISO/IEC, 2015). Każdy boks jest wyodrębnioną częścią danych, a w każdym pliku znajdują się boksy obowiązkowe oraz opcjonalne. Wszystkie dane w pliku, czyli zarówno dane multimedialne, jak i dodatkowe, zawarte są w boksach, a norma nie przewiduje istnienia innej formy

zapisu danych. Każdy boks zaczyna się nagłówkiem z informacją o jego rozmiarze oraz o typie boksu. Co istotne, dane w boksach zapisywane są w konwencji *big-endian*, czyli najbardziej znaczący bajt jako pierwszy, w odróżnieniu od popularnego formatu zapisu multimedialnych RIFF, gdzie obowiązuje kolejność odwrotna, tj. *little-endian*. Każdy z boksów może zawierać wewnątrz inne boksy, co tworzy uporządkowaną i hierarchiczną strukturę pliku. Norma ISO/IEC 14496-12 rozróżnia około 100 uporządkowanych boksów zawierających pola z informacjami (metadanymi) lub danymi multimedialnymi.

Rozmiar boksu jest liczbą całkowitą, która wskazuje na to, ile bajtów znajduje się w danym boksie. Wartość ta określa jego całkowitą zawartość, tzn. zarówno nagłówek, jak i pola z danymi oraz ewentualne inne boksy wewnątrz boksu głównego. Wartość rozmiaru boksu może być prezentowana w postaci kompaktowej albo rozszerzonej, czyli zapisana z wykorzystaniem 32 albo 64 bitów. Aktualnie typowy rozmiar boksu zapisywany jest na 32 bitach, co jest wartością wystarczającą. Jedynie boksy zawierające dużo danych multimedialnych albo zakodowaną datę i czas mogą wymagać 64 bitów.

Typ boksu identyfikuje obiekt w pliku oraz określa, jakiego rodzaju informacje są w nim zapisane. Wskazuje on, czy boks zawiera zakodowane dane multimedialne czy też metadane, a jeżeli tak, to jakiego rodzaju. Typ boksu zapisywany jest zwykle w wersji kompaktowej jako 4 znaki alfanumeryczne w celu jego łatwiejszej identyfikacji w kodzie ASCII. Wiele boksów bezpośrednio za określeniem ich typów zawiera również dodatkowe informacje, takie jak numer wersji oraz tzw. flagi. Numer wersji boks, przyjmujący zwykle wartość 0 albo 1, wskazuje najczęściej, czy rozmiar boks jest w postaci kompaktowej czy rozszerzonej. Z kolei flagi sygnalizują o stanie aktywności dodatkowych opcji lub o informacjach w obrębie boks. Przykładem jest notyfikacja o tym, że ścieżka dźwiękowa jest dostępna dla użytkownika, znajduje się w aktualnym pliku i możliwe jest jej odtworzenie. Na rysunku 1 przedstawiono ogólną strukturę pojedynczego boks w pliku zgodnym z normą ISO/IEC 14496-12 obowiązującą dla wszystkich boksów w każdym pliku zapisanym według wymienionej normy.

Format pliku zgodny z omawianą normą uwzględnia również przypadek, gdy rozmiar danych multimedialnych jest znaczny i wykracza poza zakres jednego pliku. Wówczas jeden plik zawiera jedynie metadane, a w następnych plikach zapisane są multimedia do prezentacji.

Typowy plik multimedialny zapisany według normy jest samowystarczalny i zawiera wszystkie niezbędne informacje potrzebne do odtworzenia zawartych w nim mediów. Znajdują się w nim główne boksy, które określają: typ pliku (*File Type Box*), zakodowane dane (*Media Data Box*) oraz metadane (*Movie Box*). Podstawową strukturę pliku zgodnego z normą ISO/IEC 14496-12 przedstawiono na rys. 2.

W strukturze pliku mogą znajdować się również inne główne boksy, w tym określające pofragmentowane dane multimedialne (*Movie Fragment(s) Box*) lub wolną przestrzeń w pliku (*Free Space Box*). W celu zapewnienia większej efektywności w procesie odczytu informacji norma ISO/IEC 14496-12 definiuje kolejność boksów głównych w strukturze pliku; poniżej przytoczono najbardziej istotne. Zaleca się, aby boks określający typ pliku (*File Type Box*) znajdował się na samym jego początku, przed innymi głównymi boksami. Z kolei boks z metadanymi mediów (*Movie Box*) powinien być zlokalizowany na końcu pliku. W większych boksach głównych występują dodatkowe mniejsze boksy z nagłówkami (*Header Boxes*), które muszą być umiejscowione jako pierwsze w obrębie głównego boksu. Jeśli dane multimedialne są pofragmentowane, zaleca się, aby kolejne fragmenty w jednym albo wielu plikach zostały ułożone według ich kolejności. Na rys. 3 przedstawiono uproszczoną strukturę danych w przykładowym pliku zgodnym z omawianą normą. Uproszczenie polega na prezentacji jedynie najistotniejszych boksów w celu zarysowania idei takiej struktury. Część boksów, takich jak *Movie Box*, posiada rozbudowaną hierarchię obejmującą nawet kilka poziomów w głąb.

Uwzględniając możliwość projektowania własnych obiektów, norma informuje, że w przypadku braku rozpoznania któregoś z boksów powinien on zostać zignorowany, a proces odczytu danych kontynuowany od następnego rozpoznanego boksu. Niektóre boksy, jak na przykład te związane z wolną przestrzenią (*Free Space Box*), mogą zostać nawet usunięte z pliku bez wpływu na możliwość odtworzenia zapisanych multimediiów.

Metadane znajdujące się w pliku zawierają wpisy z informacjami dla badań autentyczności nagrań związane z parametrami zapisu, sygnaturami czasowymi lub urządzeniem rejestrującym (Kajstura i in., 2017; Michalek, 2017). W przypadku plików zapisanych zgodnie z normą ISO/IEC 14496-12 najistotniejszym obiektem jest boks o nazwie *Movie Box*, w którym zawarte są metadane mediów. Obiekt ten zostanie szczegółowo omówiony w dalszej części pracy.

Celem niniejszej pracy było ustalenie, jakie informacje znajdują się w obrębie plików dźwiękowych zapisanych zgodnie z normą ISO/IEC 14496-12 wraz z oceną możliwości ich zastosowania w badaniach autentyczności nagrań. Dodatkowo opracowano własne narzędzie do automatycznej analizy struktury tego rodzaju plików i prezentacji ich właściwości.

3. Materiał do badań i metody

W celu wykonania zamierzonych badań zgromadzono i przeanalizowano możliwości rejestracji urządzeń mobilnych następujących marek i modeli:

- telefony komórkowe: LG L9, MANTA MSP 4505, Samsung Galaxy Ace 2 CM11, Samsung Galaxy Core Plus SM-G350, Samsung Galaxy Ace 3 LTE GT-S7275R, Samsung Galaxy Core Prime SM-G361F, Samsung GT-I9060 Galaxy Grand Neo, Samsung GT-C3350, Samsung Galaxy J3, Sony Xperia J1, Sony Ericsson WT19i;
- tablet ASUS Memo Pad HD7.

Za pomocą wymienionych urządzeń zarejestrowano 105 nagrań z zapisem dźwięku (tj. bez obrazu) w formatach 3GP oraz MP4(M4A). W dalszej części pracy pliki formatu MP4 ze ścieżką dźwiękową oznaczane będą jako pliki M4A. Do realizacji nagrań wykorzystano zainstalowane fabrycznie aplikacje umożliwiające rejestrację dźwięku i zapis do pliku. Nagrania te wykonano z wykorzystaniem dostępnych funkcji podczas rejestracji, tj. uruchomienia i zakończenia nagrywania oraz pauzy. Pliki z nagraniami zapisywane były zarówno w pamięci urządzeń, jak i na wymiennej karcie pamięci typu microSD. Do bezpiecznej eksploracji zawartości pamięci z nagraniami wykorzystano urządzenie uniemożliwiające modyfikację danych na tego typu nośnikach, czyli tzw. *write-blocker*. W celu dalszych analiz wykonano kopie robocze plików z nośników źródłowych na dysk twardy komputera laboratoryjnego.

Aby ocenić możliwość wykorzystania informacji zapisanych w strukturze utrwalonych plików do badań autentyczności nagrań, poddano je następującym analizom:

- właściwości plików odczytywanych za pomocą systemu operacyjnego Windows 7 Ultimate oraz dwóch programów przeznaczonych do analizy plików multimedialnych, tj. *ffprobe v.N-90893-gcae6f806a6* i *MediaInfo v.17.10*;
- na podstawie normy ISO/IEC 14496-12 i z wykorzystaniem edytorów szesnastkowych 010 Editor v.6.0.2 i HxD v.1.7.7.0 wizualizacja, odczyt i interpretacja struktury danych w utrwalonych plikach w celu identyfikacji poszczególnych obiektów (boksów);
- ustalenie zawartości informacji (metadanych) w ujawnionych boksach i ich istotności dla analizy autentyczności.

4. Wyniki

4.1. Analiza właściwości plików

Celem niniejszej części badań było ustalenie, jakie są możliwości odczytu i interpretacji parametrów plików dźwiękowych 3GP i M4A za pomocą standardowo stosowanych narzędzi. Na wstępie wykonano analizę informacji prezentowanych przez system operacyjny Windows 7 Ultimate. Ustalono, że system ten prezentuje podstawowe informacje o obu typach plików: nazwę z rozszerzeniem, rozmiar, sygnatury czasowe, jak również

przybliżone wartości przepływności bitowej oraz czasu trwania nagrania. Dla plików w formacie 3GP prezentuje dodatkowo liczbę kanałów i przybliżoną wartość częstotliwości próbkowania dźwięku, klasyfikuje je również jako pliki wideo.

Do odczytu informacji o plikach multimedialnych służą także dedykowane programy, takie jak *ffprobe* i *MediaInfo*. Są to często wykorzystywane narzędzia, które dostarczają bardziej precyzyjnych informacji o tego rodzaju plikach. Rozpoznają one wiele formatów zapisu multimedialnych od popularnych po rzadko spotykane. Oba wymienione programy oprócz informacji podstawowych, jak dla systemu Windows, prezentują również szczegółowe parametry zapisu dźwięku w plikach 3GP i M4A: format, profil formatu, kodowanie, częstotliwość próbkowania, przepływność, czas trwania oraz sygnatury czasowe ogólne dla pliku i danych dźwiękowych. Na podstawie zestawienia informacji odczytanych za pomocą wymienionych wyżej narzędzi ustalono, że zarejestrowane pliki dźwiękowe odznaczają się następującymi istotnymi parametrami:

- pliki w formacie 3GP mają rozszerzenia *.3gpp, natomiast M4A rozszerzenia *.m4a;
- dźwięk we wszystkich plikach utrwalony został w postaci skompresowanej: w plikach 3GP zastosowano kodowanie AAC-LC albo AMR-NB, natomiast w plikach M4A kodowanie AAC-LC;
- przepływności bitowe dla plików 3GP zawierały się w przedziale 32–156 kb/s (AAC-LC) oraz 12 kb/s (AMR-NB), z kolei dla plików M4A w granicach 56–128 kb/s;
- częstotliwości próbkowania dla plików 3GP wynosiły 44100 Hz (AAC-LC) albo 8000 Hz (AMR-NB), a dla plików M4A w przedziale 16000–48000 Hz,
- w zależności od wybranych opcji programów *ffprobe* i *MediaInfo* czas trwania nagrań prezentowany jest z różną precyzją; co ważne, sygnatury czasowe odczytywane są przez te programy z metadanych w pliku (w *MediaInfo* utworzenia oraz modyfikacji, w *ffprobe* jedynie utworzenia).

4.2. Analiza informacji znajdujących się w metadanych

Norma ISO/IEC 14496-12 opisuje architekturę plików multimedialnych oraz definiuje znaczenie i strukturę znajdujących się w nich obiektów zwanych boksami. W niniejszej pracy dokonano przeglądu oraz opisu zgodnych z normą boksów znajdujących się w zarejestrowanych plikach 3GP i M4A zawierających istotne metadane dla badań autentyczności tego rodzaju nagrań. Dodatkowo, na rycinach przedstawiono przykłady w postaci wizualizacji boksów w kodzie heksadecymalnym i ASCII. Każdy istotny boks oraz jego pola zostały z detalami opisane w celu wykorzystania w konstruowanym programie

komputerowym do automatycznej analizy plików opartych na normie ISO/IEC 14496-12.

File Type Box

Pliki multimedialne zgodne z normą ISO/IEC 14496-12 mogą być kompatybilne z jedną lub większą liczbą szczegółowych specyfikacji technicznych. Boks *File Type* jest jednym z głównych i obowiązkowych obiektów w pliku i powinien być usytuowany na samym początku jego struktury, przed innymi głównymi boksami. Określa on typ pliku i zawiera dodatkowe informacje o obsługujących go i kompatybilnych specyfikacjach. Informacje w tym boksie mogą wskazywać na rodzaj zastosowanego kodowania, rozszerzenie pliku, wymagania co do obsługi konkretnych obiektów lub na jego ogólne przeznaczenie. Omawiany boks zawiera informacje istotne dla aplikacji odtwarzających w celu uzyskania najlepszej jakości odczytu danych, jak również ważne w procesie inspekcji pliku i usuwania ewentualnych błędów.

Obiekt *File Type* w polu *Box Type* nagłówka zawiera ciąg znaków *ftyp*, który identyfikuje ten boks w strukturze całego pliku. Omawiany boks zawiera następujące pola:

- *major_brand*: identyfikuje najlepszą specyfikację do obsługi pliku,
- *minor_version*: wersja tej specyfikacji,
- *compatible_brands*: specyfikacje kompatybilne.

Norma ISO/IEC 14496-12 podaje przykłady kilkunastu specyfikacji, które mogą zostać zaimplementowane w obiekcie *File Type*, w tym: *ISO Base Media File* w wersji 1 (oznaczona wpisem *isom*) i w wersjach wyższych (przykładowo – w wersji 2 jako *iso2*), AVC (oznaczony *avc1*) i MPEG w. 7 (*mp71*). Istnieje oczywiście możliwość implementacji innych specyfikacji pod warunkiem ich zdefiniowania. Wyższe wersje danych specyfikacji wprowadzają obsługę dodatkowych funkcji lub boksów.

W strukturze wszystkich zgromadzonych plików 3GP i M4A zidentyfikowano boks *File Type*, który znajdował się na początku danych.

Dla plików 3GP pola *major_brand* wskazywały na specyfikacje 3GPP w. 4 (wpis *3gp4*), natomiast *compatible_brands* na *ISO Base Media File* w. 1 oraz 3GPP w. 4.

Dla plików M4A pola *major_brand* wskazywały na specyfikacje MP4 w. 2 (wpis: *mp42*) i 3GPP w. 4, a *compatible_brands* na *ISO Base Media File* w. 1, 3GPP w. 4 i MP4 w. 2.

W przypadku zdecydowanej większości zgromadzonych plików 3GP i M4A rozmiar całego boksów *File Type* wynosił 24 bajty i zawierał jeden wpis *major_brand* oraz dwa wpisy *compatible_brands*. Z kolei pliki M4A zarejestrowane za pomocą telefonu Samsung GT-C3350 odznaczały się rozmiarem 28 bajtów i trzema wpisami *compatible_brands*, co jest pewną cechą wyróżniającą

ten model telefonu spośród pozostałych i dodatkową informacją w przypadku analizy autentyczności. Żaden z analizowanych boksów *File Type* nie miał w swojej strukturze boksów wewnętrznych. Na rysunku 4 przedstawiono fragment struktury danych w pliku formatu 3GP z zaznaczonym boksem *File Type*.

Pliki zgodne ze starszymi specyfikacjami mogą nie zawierać boksów *File Type*. Bieżąca norma przewiduje taki przypadek i wskazuje, aby odczytując takie pliki, przyjmować wartość *major_brand* i *compatible_brand* jako *mp41*, natomiast wartość *minor_version* jako 0.

Zakodowane multimedia, *Media Data Box*

Zakodowane dane multimedialne zawarte są w boksie *Media Data*. Jest to jeden z głównych boksów w strukturze pliku, choć nie jest obowiązkowy. Zgodnie z normą może bowiem istnieć plik zawierający jedynie metadane, natomiast powiązane z nimi multimedia zawarte są w osobnych plikach. Dla metadanych opisujących dane multimedialne obowiązujące jest adresowanie bezwzględne, tj. liczone jako offset od początku danych w całym pliku. Możliwe jest również, że jeden plik będzie zawierał kilka boksów *Media Data*.

Analiza struktury danych zarejestrowanych plików 3GP i M4A wykazała, że w każdym z nich znajdował się dokładnie jeden boks z multimediami. Boks *Media Data* w nagłówku o rozmiarze 8 bajtów zawiera ciąg znaków *mdat* identyfikujący go w obrębie pliku. Nagłówek zawiera również informacje o rozmiarze tego boksów, przy czym wartość ta zależy od rozmiaru danych multimedialnych. Omawiany boks w zarejestrowanych plikach dźwiękowych nie miał w swojej strukturze innych, wewnętrznych boksów. Dla analizowanych nagrań był to najobszerniejszy obiekt w całym pliku multimedialnym.

Struktura boksów *Media Data* nie jest skomplikowana w porównaniu z innymi obiektami, pozwala jednak na dość istotną analizę. Znany jest rozmiar nagłówka i zadeklarowany w nim rozmiar całego boksów *mdat*. Możliwe jest również odczytanie wprost z zapisu heksadecymentalnego rzeczywistego rozmiaru samych danych multimedialnych i porównanie z wartością podaną w nagłówku. Niezgodność tych dwóch wartości mogłaby wynikać na przykład z usunięcia części danych. Zgodnie z omówioną wyżej strukturą pliku ISO/IEC możliwe jest również zbadanie, czy bezpośrednio po boksie *Media Data* znajduje się nagłówek innego boksów, co wskazywałoby na prawidłową strukturę. Stanowi to dodatkową weryfikację rozmiaru zadeklarowanych danych multimedialnych. Na podstawie analizy heksadecymalnej zgromadzonych plików 3GP i M4A ustalono, że bezpośrednio za danymi multimedialnymi znajdował się nagłówek boksów z metadanymi. Na rys. 5 przedstawiono fragment struktury pliku M4A z zaznaczonym nagłówkiem boksów *mdat*.

Metadane w pliku, *Movie Box*

Najistotniejszym obiektem dla badań autentyczności nagrań zapisanych w plikach według omawianej normy jest boks o nazwie *Movie Box*. Jest to jeden z głównych i obowiązkowych obiektów w strukturze pliku, w którym znajdują się metadane dla całej prezentacji danych. Według zaleceń normy boks z metadanymi powinien być zlokalizowany na końcu całego pliku.

Na podstawie analizy struktury zgromadzonych plików 3GP i M4A stwierdzono, że *Movie Box* jest bardzo rozbudowanym obiektem o hierarchicznej i wielopoziomowej budowie zawierającym liczne wewnętrzne boksy. Na jego początku znajduje się typowy nagłówek boksów o rozmiarze 8 bajtów wraz z ciągiem znaków *moov* jako identyfikatorem. W nagłówku tym występuje również pole z rozmiarem całego boksów, przy czym wartość ta nie jest stała. Zależy ona od wielu czynników, przede wszystkim od liczby boksów ustalonej przez projektanta, szczegółowości opisów, liczby ścieżek, a nawet od rozmiaru danych multimedialnych. W odniesieniu do pozostałych głównych boksów jest to zwykle jeden z najobszerniejszych obiektów. Zawiera on średnio od kilkudziesięciu do kilkuset kilobajtów, a w przypadku dłuższych nagrań nawet ponad megabajt informacji (ustalono na podstawie nagrań z typowym kodowaniem dźwięku AAC-LC, próbkowaniem 44100 Hz trwających od kilkudziesięciu sekund do 2 godzin). W każdym pliku 3GP i M4A stwierdzono występowanie dokładnie jednego obiektu *Movie Box*.

W obrębie boksów *Movie Box* znajdują się wewnętrzne boksy, zarówno obowiązkowe, jak i fakultatywne. Znaczenie i zawartość najistotniejszych z nich zostaną omówione poniżej.

Nagłówek boksów z metadanymi, *Movie Header*

Z uwagi na obszerność i złożoność obiektu *Movie Box* zawiera on dodatkowy nagłówek w postaci wewnętrznego boksów *Movie Header*. W strukturze każdego analizowanego pliku 3GP i M4A stwierdzono jeden tego typu obiekt. Znajduje się on bezpośrednio za 8-bajtowym nagłówkiem na początku głównego boksów, tj. *Movie Box*, i jest jego obowiązkowym elementem. W nagłówku *Movie Header* znajduje się wpis *mvhd* który jest jego identyfikatorem. W boksie tym znajdują się ogólne informacje dotyczące całej prezentacji danych multimedialnych. Na podstawie normy oraz zgromadzonych plików 3GP i M4A ustalono, że zawiera on następujące istotne pola:

- *version*: wersja boksów, może przyjmować wartość 0 albo 1; w przypadku wartości 0 opisane niżej pola *creation_time*, *modification_time* oraz *duration* mają rozmiar 32 bitów, natomiast dla wartości 1 mają one rozmiar 64 bitów;

- *creation_time* i *modification_time*: wartość związana, odpowiednio, z utworzeniem i modyfikacją prezentacji multimedialnych; liczona jest co 1 sekundę od północy 1 stycznia 1904 roku i przedstawiana w formacie czasu *Universal Time Coordinated* (UTC);
- *timescale*: liczba jednostek czasu w ciągu 1 sekundy;
- *duration*: czas trwania prezentacji w sekundach dla zdefiniowanej skali czasu równy czasowi trwania najdłuższej ścieżki w całej prezentacji,
- *next_track_ID*: identyfikator kolejnej ścieżki w prezentacji.

Analiza wszystkich zarejestrowanych plików z nagraniami pozwoliła ustalić dodatkowo, że wartości kolejnych pól równały się: *version* – 0, *timescale* – 1000 (tzn. czas liczony co 1 milisekundę), *next_track_ID* – 2, a parametr *duration* równy był czasowi trwania całego nagrania.

Na szczególną uwagę zasługują pola *creation_time* i *modification_time*. Zawierają one liczbę sekund zapisaną w reprezentacji szesnastkowej, którą należy zdekodować do systemu dziesiętnego, sformatować do postaci daty i godziny oraz obliczyć różnicę od północy 1 stycznia 1904 roku. W ten sposób wyznaczyć można datę i godzinę utworzenia oraz modyfikacji. Na podstawie badań ustalono, że tak obliczone pary czasów utworzenia i modyfikacji dla każdego pliku 3GP i M4A były takie same i odnosiły się do czasu zakończenia rejestracji. W badaniach autentyczności należy również uwzględnić różnicę między czasem środkowoeuropejskim a czasem UTC, która wynosi +1 albo +2 godziny w zależności od czasu zimowego albo letniego. Na rys. 6 przedstawiono strukturę danych w pliku M4A z zaznaczonym fragmentem *Movie Header* zawierającym wpisy, na podstawie których można wyznaczyć omawiane sygnatury czasowe dla prezentacji.

Dla analizowanych plików ustalono, że obliczone sygnatury czasowe zależą od ustawienia zegara systemowego urządzenia, przy czym menu badanych urządzeń umożliwiało zmianę tego parametru przez użytkownika. Domyślnie jednak urządzenia mobilne, takie jak telefony komórkowe oraz tablety i smartwatche z funkcją telefonu, mają włączoną opcję automatycznej aktualizacji czasu dostarczanego od operatora sieci. Jeżeli są w zasięgu sieci, wówczas ustawiony czas powinien być aktualny, także ten zakodowany w boksie *Movie Header*. Jeżeli jest to możliwe, sygnatury czasowe wyznaczone na podstawie wskazanych wyżej metadanych powinny być weryfikowane innymi sposobami, na przykład metodą wykorzystującą sygnał przydzwięku sieciowego.

Czasy obliczone na podstawie wpisów w polach *creation_time* i *modification_time* są niezależne od wskazań systemu operacyjnego Windows i nie zmieniają się po wykonaniu kopii pliku na inny nośnik danych, co jest kluczowe dla analizy autentyczności. Często bowiem do badań przekazywane są jedynie kopie nagrań na wy-

miennych nośnikach, które odznaczają się już zmodyfikowanymi sygnaturami czasowymi systemu plikowego.

Ścieżki w pliku multimedialnym, *Track Box*

Pliki multimedialne mogą zawierać wiele ścieżek, takich jak wideo lub dźwiękowe. Ich opis znajduje się w obiektach *Track Box*, a dla każdej pojedynczej ścieżki przeznaczony jest osobny boks tego typu. Jest to jeden z podstawowych i obowiązkowych obiektów w głównym boksie *Movie Box*. W każdym analizowanym pliku 3GP i M4A stwierdzono dokładnie jedną ścieżkę dźwiękową, zapisaną w jednym obiekcie *Track Box*. Identyfikatorem tego boksu jest ciąg znaków *trak*, znajdujący się w jego nagłówku.

Obiekt *Track Box* posiada dodatkowy nagłówek *Track Header Box* jako wewnętrzny, pojedynczy i obowiązkowy obiekt o identyfikatorze *tkhd*. Nagłówek ten zawiera ogólne metadane dla zapisanej ścieżki i znajdują się w nim następujące istotne pola:

- *version*: wartość 0 albo 1; wskazuje rozmiar sygnatur czasowych i czasu trwania ścieżki na, odpowiednio, 32 albo 64 bity;
- *creation_time* i *modification_time*: wartość związana z utworzeniem i modyfikacją ścieżki; podobnie jak w *Movie Header* czas liczony jest co 1 sekundę od północy 1 stycznia 1904 roku i przedstawiany jako *Universal Time Coordinated* (UTC);
- *track_ID*: wartość unikalna i identyfikująca ścieżkę w pliku;
- *duration*: czas trwania ścieżki dla zdefiniowanej skali czasu w *Movie Header*;
- *flags*: pole zawierające tzw. flagi, które definiują stan ścieżki, czyli jej dostępność albo zablokowanie dla użytkownika; określają, czy ścieżka jest używana w prezentacji i możliwa do odtworzenia.

Na podstawie zgromadzonych nagrań 3GP i M4A ustalono, że sygnatury czasowe utworzenia i modyfikacji ścieżki oraz czas jej trwania były zgodne z odpowiadającymi wartościami dla całej prezentacji. Czas trwania ścieżki dla każdego analizowanego pliku wyrażony był w milisekundach. Wartości kolejnych pól wyniosły: *version* – 0, *track_ID* – 1 (tzn. pierwsza ścieżka w pliku), natomiast *flags* – 7 (wartość typowa jako suma kolejnych flag), co wskazywało, że ścieżka jest: dostępna, używana w prezentacji i możliwa do odtworzenia.

Informacje o mediach w ścieżce, *Media Box*

Zagłębiając się coraz bardziej w hierarchii metadanych *Movie Box*, należy wyróżnić boks zawierający informacje związane z mediami zapisanymi w ścieżce, czyli *Media Box*. Jest to boks obowiązkowy i wewnętrzny w obrębie *Track Box*. W każdym pliku 3GP i M4A

występował jeden taki obiekt, który odznaczał się ciągiem znaków *mdia* znajdującym się w jego nagłówku.

Podobnie jak omawiane poprzednio obiekty, *Media Box* posiada swój wewnętrzny nagłówek o nazwie *Media Header Box* i identyfikatorze *mdhd* z ogólnymi informacjami o mediach w ścieżce. Ustalono, że zawiera on następujące istotne pola:

- *version*: wskazuje rozmiar sygnatur i czas trwania mediów (0 dla 32, 1 dla 64 bitów);
- *creation_time* i *modification_time*: sygnatury utworzenia i modyfikacji mediów, których zasada wyznaczania jest taka sama jak dla ścieżki w pliku;
- *timescale*: wartość podstawy czasu dla zapisanych mediów;
- *duration*: czas trwania mediów w ścieżce w odniesieniu do parametru *timescale*;
- *language*: język zastosowany dla mediów.

Sygnatury czasowe utworzenia i modyfikacji oraz czas trwania mediów w analizowanych plikach dźwiękowych zgodne były z odpowiednimi sygnaturami i czasem trwania dla ścieżek. Parametr *timescale*, pomimo zbieżnej nazwy z parametrem w *Movie Header*, równał się w tym przypadku z częstotliwością próbkowania dźwięku. Dla zarejestrowanych nagrań parametr ten wynosił od 8000 do 48000. Informacja w polu *language* zapisywana jest za pomocą trzech znaków alfanumerycznych, których znaczenie opisano w normie ISO 639-2/T. W zgromadzonych plikach dźwiękowych 3GP i M4A wartość w tym polu zawsze wskazywała na język angielski pomimo ustawienia polskiego menu w badanych urządzeniach.

Wewnątrz obiektu *Media Box* znajduje się boks o nazwie *Handler Reference Box*, który wskazuje na typ ścieżki z mediami, a tym samym na sposób jej odtwarzania. Identyfikowany jest na podstawie ciągu znaków *hdlr* i posiada następujące istotne pola:

- *handler_type*: oznacza typ ścieżki z zapisanymi mediami;
- *name*: słowny opis typu ścieżki.

Zarejestrowane do badań pliki 3GP i M4A zawierały ścieżki dźwiękowe i odznaczały się wpisami: w polu *handler_type* – *sound*, natomiast w polu *name* – *SoundHandle*.

W obrębie *Media Box* znajduje się również obiekt *Media Information Box* identyfikowany ciągiem znaków *minf*, który zawiera dodatkowe wewnętrzne boksy z informacjami ważnymi dla analizy autentyczności nagrań. W obiekcie *Sample Table Box* o identyfikatorze *stbl* znajdują się boksy oznaczone jako *Sample Description Table* i *Bit Rate Box*. Struktura pierwszego z nich zależna jest od typu ścieżki (tj. od wartości *handler_type* opisanej wyżej) i zawiera pole *format* z informacjami o formacie zapisu multimedialnych. Dla zgromadzonych nagrań dźwiękowych boks ten identyfikowany był ciągiem znaków *stsd*, a pole *format* obejmowało wpisy: dla plików 3GP

– *mp4a* (MP4 dla audio, profil kodowania AAC-LC) albo *samr* (*Adaptive Multi-Rate* w wersji *Narrowband*), natomiast dla plików M4A wyłącznie *mp4a*. Jak widać, dźwięk w analizowanych nagraniach zapisywany był jedynie w postaci kompresji stratnej. Drugi z obiektów to *Bit Rate Box* identyfikowany ciągiem znaków *brtt*. Zawiera on pola *maxBitrate* i *avgBitrate*, które prezentują maksymalne i średnie wartości przepływności bitowej w bitach na sekundę dla strumienia danych. Obiekt *Sample Table Box* zawiera również informacje o indeksach czasu dla próbek zapisanych w ścieżce i różnicy aktualnego indeksu względem początku ścieżki. W boksie nazwanym *Decoding Time to Sample Box* o identyfikatorze *stts* zapisana jest w postaci kompaktowej tabela z przypisanymi indeksami czasu do numerów próbek. Według normy pojęcie próbki oznacza pojedynczą sekcję zakodowanych danych (ramkę), której można przypisać indeks czasu. Analiza zgromadzonych nagrań 3GP i M4A wykazała, że boks ten zawiera przeważnie dwie pary wpisów: pierwsza z numerem i indeksem dla pierwszej próbki, druga z podanymi wartościami dla próbki ostatniej. Na podstawie liczby próbek i ich indeksów w czasie można wyznaczyć czas trwania wszystkich próbek. Obliczona wartość powinna być zgodna z czasem trwania mediów i pozwala na dodatkową weryfikację z parametrami zapisanymi w strukturze pliku. Zgodność taką ustalono dla zarejestrowanych plików 3GP i M4A. Przykład parametrów zapisanych w boksie *stts* przedstawiono na rys. 7.

Wolna przestrzeń, *Free Space Box*

Norma ISO/IEC 14496-12 definiuje także obiekt o nazwie *Free Space Box*, który może być jednym z głównych boksov w hierarchii struktury pliku lub boksem wewnętrznym. Nie zawiera on istotnych informacji i może być pomijany w trakcie odczytu zawartości prezentacji. Możliwe jest nawet jego usunięcie z pliku bez wpływu na zapisane multimedia. Jego budowa jest nieskomplikowana i zawiera nagłówek z identyfikatorem *free* oraz dane. Analiza zgromadzonych plików pozwoliła jednak na spostrzeżenie, że ten nieobowiązkowy boks występuje niezbyt często. Zauważono go jedynie w plikach 3GP zarejestrowanych urządzeniami Asus, LG, Manta i Samsung Galaxy Ace 2. Zlokalizowany był pomiędzy boksem określającym typ pliku i boksem z multimediami, a jego dane stanowiły ciąg bajtów wypełnionych zerami. Pomimo nieznaczącej zawartości *Free Space Box* stanowi pewną cechę wyróżniającą dla wymienionych urządzeń, co może być dodatkową informacją w procesie badania autentyczności.

W tabeli 1 zestawiono kluczowe metadane, które odczytano ze struktury analizowanych plików dźwiękowych 3GP i M4A z uwzględnieniem opisanych wyżej wyników badań.

4.3. Oprogramowanie do automatycznej analizy struktury plików i metadanych

Na podstawie otrzymanych wyników badań struktury plików 3GP i M4A oraz normy ISO/IEC 14496-12 opracowano algorytmy do automatycznej analizy tego typu plików. Dokonano szczegółowej interpretacji pól i metadanych wszystkich opisanych w pracy boksów oraz stworzono bazę danych z opisem znaczenia każdego z nich na potrzeby tworzonego oprogramowania. Z wykorzystaniem środowiska obliczeniowego MATLAB utworzono program komputerowy, którego ogólna zasada działania jest następująca:

- 1) import całości pliku do przestrzeni roboczej programu,
- 2) ustalenie typu zapisanych ścieżek (wideo, dźwiękowa lub pomocnicze),
- 3) weryfikacja rozmiaru zadeklarowanych względem obliczonych danych multimedialnych,
- 4) analiza struktury zaimportowanego pliku i detekcja kolejnych boksów na podstawie ich identyfikatorów,
- 5) odczyt danych oraz metadanych z boksów w kodzie heksadecymalnym i ASCII,
- 6) określenie wartości pól i parametrów z wykorzystaniem wbudowanej bazy danych,
- 7) weryfikacja integralności struktury pliku oraz detekcja ewentualnych zaburzeń lub braku istotnych boksów,
- 8) wygenerowanie raportu z wykonanej analizy pliku zawierającego opis boksów i ich metadanych oraz wyekstrahowane parametry.

Fragment przykładowego raportu wygenerowanego na podstawie analizy metadanych w pliku M4A za pomocą opracowanego oprogramowania przedstawiono na rys. 8.

Na rys. 9 przedstawiono fragment raportu na podstawie analizy metadanych w pliku M4A, w którym uszkodzony został jeden z głównych i obowiązkowych obiektów, tj. *Movie Box*.

Podjęcie prac nad własnym oprogramowaniem umotywowane było potrzebą stworzenia narzędzia dedykowanego do analizy autentyczności nagrań, które zapewni wiarygodną analizę struktury plików i boksów wraz z precyzyjnym nazewnictwem parametrów zgodnym z zasadami normy ISO/IEC. Z uwagi na elastyczność struktury tego rodzaju plików projektanci urządzeń i aplikacji mają możliwość umieszczania w strukturze nowych boksów zawierających multimedia lub metadane. Wraz z analizą kolejnych dowodowych i testowych nagrań istnieje możliwość łatwego uaktualniania bazy danych o nowe rozpoznane boksy w celu ich wykorzystania w badaniach autentyczności.

5. Wnioski

Badania opisane w niniejszej pracy pozwoliły na pozyskanie i usystematyzowanie wiedzy o strukturze plików multimedialnych zapisanych zgodnie z normą ISO/IEC 14496-12. Do wykonania tych badań przystąpiono ze względu na zwiększającą się liczbę nagrań dowodowych poddawanych analizie autentyczności, które rejestrowane są w formatach bazujących na wymienionej normie, takich jak 3GP i MP4(M4A). Na podstawie zgromadzonych nagrań i standardu ISO/IEC ustalono, że struktura plików tego rodzaju składa się z boksów, wewnątrz których zapisane są multimedia oraz metadane. Analiza i ocena tych elementów pozwoliła wykazać, że w ich obrębie znajdują się istotne informacje, które można wykorzystać w badaniach autentyczności nagrań. Pozwalają one na ustalenie poprawności struktury pliku, parametrów zapisu multimedii i sygnatur czasowych rejestracji. Jest to istotne, gdyż analiza struktury plików jest aktualnie jedną z podstawowych metod wykorzystywanych w badaniach autentyczności nagrań. Na podstawie uzyskanych wyników wykonano oprogramowanie do automatycznej analizy struktury plików zgodnych z normą ISO/IEC 14496-12, które pracuje w środowisku obliczeniowym MATLAB. Wykonuje ono analizę integralności danych i wyznacza poprawne parametry oraz ewentualne zaburzenia struktury pliku. Stanowi narzędzie wspomagające w badaniach autentyczności nagrań wykonywanych w związku z działalnością opiniodawczą Instytutu.