



# PROPERTIES OF RECORDINGS AND AUDIO FILES SAVED IN AMR FORMAT AND AN ASSESSMENT OF THE POSSIBILITY OF APPLYING THEM IN AUTHENTICITY EXAMINATIONS

Marcin MICHAŁEK

*Institute of Forensic Research, Kraków, Poland*

## Abstract

One of the commonly used formats for recording audio on mobile devices is Adaptive Multi-Rate (AMR). It was primarily designed to transmit and store speech, although it can also be used to record other acoustic signals. AMR file format is mainly used for recording audio on mobile phones with a voice recorder function. It can also be used for storing so-called voice messages, offered as a service by mobile telephony operators. Recordings originating from mobile phone memories, including in AMR format, are currently a significant part of all analysed evidence recordings. That is why it is important to analyse the authenticity of AMR audio recordings. The purpose of this work was to analyse recordings and audio files stored in AMR file format and to study their application in authenticity examinations. Another goal was to create our own software to automatically analyse the file structure of such recordings. The study was based on AMR audio recordings which were made using different brands and models of mobile phones. The obtained results allowed us to establish that recording in the seemingly uncomplicated AMR file format enables the authenticity of such recordings to be investigated. It is also possible to use well-known methods such as the Electrical Network Criterion and file properties and structure in such investigations. An algorithm and software for the automatic analysis of the integrity of AMR audio files in the MATLAB computing environment were also developed.

## Key words

AMR; Authenticity analysis; Digital recording; File structure.

*Received 24 July 2017; accepted 19 September 2017*

## 1. Introduction

Adaptive Multi-Rate (AMR) codec was standardized by the European Telecommunications Standards Institute (ETSI) and has been developed since 1999 by the 3rd Generation Partnership Project (3GPP). It is available in two variants, Adaptive Multi-Rate Narrowband, also referred to as AMR-NB, and Adaptive Multi-Rate Wideband, AMR-WB (3GPP, 2011a; ETSI, 2002). AMR and AMR-WB were originally designed for application in radio communication systems. Due to the flexibility of this codec and the possibility of its adaptation to connections of varying quality, it has also found application in GSM and UMTS mobile

cellular telephony (3GPP, 2011b; Sjöberg, Westerlund, Lakaniemi, Xie, 2007). Apart from application in mobile communication systems, AMR is also used for transmitting speech in audio and video conferences and also in satellite and VoIP communication. It ensures acceptable speech signal quality even for weak connection parameters (Bhatt, Kosta, 2011; Sjöberg et al., 2007). This format is also applied to speech playback in audiobooks and in voice messages.

AMR is also a file format for audio recordings, which uses the narrow band version of the above mentioned codec (3GPP, 2011a; Sjöberg et al., 2007). Audio recordings in this format are recorded mainly with the use of mobile phones, but also by other popular

mobile devices, e.g., tablets (Luo, Yang, Huang, 2015; Michalek, 2016). AMR enables you to save recordings onto audio files of differing quality at a bit rate of from 4.75 to 12.2 kb/s (3GPP, 2011a). This format can also be applied to storing sound inside the structure of so-called multimedia containers, in which video or audio recordings are saved. Examples are files in 3GP and MP4 format, which are very popular in modern mobile devices, especially mobile phones (3GPP, 2010; Gloe, Fischer, Kirchner, 2014).

This paper focuses on the properties of audio recordings and files stored in AMR-NB format and the possibilities of using them in authenticity examinations. The study was conducted using audio recordings that were recorded by mobile phones with a voice recorder function and saved in so-called voice mailboxes of mobile network operators.

Currently, authenticity examinations of digital recordings require the concatenation of results and conclusions from various audio research methods, as well as – frequently – computer forensic methods (Kajstura, Michalek, Trawińska, 2017). The results of the performed research show that making recordings in the relatively uncomplicated AMR format allows the use of recognised and applied methods for authentication examinations, such as analysis of the electrical network frequency as well as file properties and file structure (Hong, Yin, 2013; Grigoras, 2005; Kajstura et al., 2017).

The analyses described in this article were conducted as part of a research project carried out at the Institute of Forensic Research in Krakow.

## 2. Audio coding in AMR format files

AMR format files do not contain coded sound sample after sample, but in the form of coefficients describing successive fragments of recording. This codec uses a linear prediction algorithm called Algebraic Codebook Excitation Linear Prediction (ACELP), which is a modification of the popular Linear Predictive Coding (LPC) algorithm used for coding and describing a speech signal (3GPP, 2011b; Luo et al., 2015).

Before the actual process of audio coding using AMR is carried out, initial preparation of the audio signal occurs, which consists of conversion of the analogue signal into a digital form with the following parameters: 13 bit resolution, sampling frequency 8000 Hz and Pulse Code Modulation (PCM) coding. The signal is also subjected to high-pass filtering with a cut-off frequency of 80 Hz (3GPP, 2011b).

The AMR codec works in such a way that it divides the entire recording into shorter fragments – so-called frames – of 20 ms duration. For a sampling frequency of 8000 Hz, an individual frame contains 160 samples. In each frame, the speech signal is analysed and parameters of the ACELP model are determined: LPC coefficients, adaptive and fixed codebooks as well as gains. In order to reproduce the content of the recording, the decoder, on the basis of parameters of the ACELP model, performs a process of synthesis of the coded signal (3GPP, 2011a; ETSI, 2002).

The following is a brief description of the principles of operation of the AMR encoder on the basis of (3GPP, 2011a, 2011b; ETSI, 2002; Sjoberg et al., 2007). The process of recording sound in a file using the AMR codec can be divided into several main stages. In the first stage, a short-term analysis is performed and coefficients of linear prediction (LPC) of the 10<sup>th</sup> order are calculated using the Levinson-Durbin algorithm, according to the formula:

$$H(z) = \frac{1}{1 + \sum_{i=1}^m \hat{a}_i z^{-i}} \quad (1)$$

where:  $\hat{a}_i$  are the linear prediction coefficients, and  $m = 10$  is the order of prediction.

Formula (1) enables short-term analysis of the signal in which the vocal tract is modelled. At this stage, long-term analysis is also carried out and parameters of the pitch function of the vocal tract are calculated with the aim of later synthesis of the speech signal by means of the following formula:

$$\frac{1}{B(z)} = \frac{1}{1 - g_p z^{-T}} \quad (2)$$

where:  $T$  is the pitch delay and  $g$  is the pitch gain.

A so-called Adaptive Codebook – described in greater detail below – is created on the basis of long-term analysis.

At the next coding stage, the already calculated LPC coefficients are converted into Line Spectral Pairs (LSP) for further quantization. This conversion is conducted on the basis of the decomposition of function  $A(z)$  into polynomials  $F_1'$  and  $F_2'$  according to the formula:

$$F_1'(z) = A(z) + z^{-11}A(z^{-1}) \quad (3)$$

$$\text{and } F_2'(z) = A(z) - z^{-11}A(z^{-1})$$

Next the determined LSP coefficients are subjected to quantization in the frequency domain in accordance with the formula:

$$f_i = \frac{f_s}{2\pi} \arccos(q_i) \quad (4)$$

where:  $i=1, \dots, 10$ ,  $f_i$  are the Line Spectral Frequencies (LSF) in the range  $[0, 4000]$  Hz corresponding to LSP coefficients,  $f_s$  is the sampling frequency of 8000 Hz, whilst  $q_i$  are the LSP coefficients in the cosine domain. Coefficients quantized in the above way are then subjected to linear interpolation, and then transformed back into quantized LPC coefficients.

In the next stage, a single frame containing 160 samples is divided into four equal parts – so-called subframes – with 40 samples each, and each lasting 5 ms. Depending on the bit rate of the AMR Codec, quantized and non-quantized LPC coefficients and interpolated LSP coefficients are used in various subframes. Parameters of the Adaptive Codebook and the Fixed Codebook are determined on the basis of subframes. The Adaptive Codebook is determined on the basis of long-term analysis and contains excitation vectors of variable values with the aim of adaptation to parameters of the speech signal for each subframe. However, the Fixed Codebook contains excitation vectors with fixed values. At the last stage of coding of AMR, techniques are used to determine pitch and their gains as well as intervals between pitch for the analysed frames (Open-loop Pitch Search and Closed-loop Pitch Analysis) in order to determine the most optimal parameters of the encoded speech signal.

The saving of the recording into the file occurs in such a way that for each frame of 20 ms, the coder generates 95, 103, 118, 134, 148, 159, 204 or 244 bits for a bit rate of 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2 or 12.2 kb/s, respectively. After encoding a single frame, the memory of the coder is updated with the aim of calculating parameters for the next frame. The structure of each AMR file consists of its header, followed by an ordered collection of frames containing successive coded fragments of the recording (3GPP, 2011b; Sjoberg et al., 2007).

### 3. Research material and applied methods

In order to conduct the intended research, recording functions on 25 mobile phones of the following brands and models were analysed:

- BlackBerry: 8520 and 9320;
- HTC Wildfire;
- LG 9;
- MaxCom MM461BB;
- Nokia: 2700, 3100c, 6103, 6230, 6300, 6500c, C1-01, C2-02, E50, E52-1, E65, and N95;

- Samsung: Galaxy Ace III, GT-C3350 and Solid B2100;
- Siemens C75;
- Sony Ericsson: W595S and WT19i.

181 test recordings were made in AMR format using the above mentioned phones. The voice recorder function and message recording service in cellular operators' voicemail boxes were used for this purpose, and messages were then saved into the memory of telephones in AMR format files. During recording, the available functions were made use of – i.e., the start, stop and pause functions. They were recorded in various acoustic conditions – in closed rooms and in open spaces – and contain a range of sounds and intentional noises over a reasonably broad range of frequencies. After being saved in the phone memory, each of the test recordings was copied onto the hard disc of the laboratory computer, or a memory image was made for the device.

In order to assess the possibility of using parameters of recordings and files stored in AMR format in authenticity examinations, the test recordings were subjected to analyses of the following:

- time and frequency characteristics with simultaneous auditory analysis;
- electrical network frequency signal – a special case of analysis of the characteristics of the recording;
- properties of files that are read by the Windows operating system and by tools designated for analysis of multimedia files;
- structure of data saved in AMR format files, using programmes for visualisation and editing in hexadecimal and ASCII code.

## 4. Results

### 4.1. Analysis of characteristics and auditory analysis

Analysis of time and frequency characteristics combined with auditory analysis is a research method applied mainly to the assessment of the continuity of a recording (Błasikiewicz, Miściuk, Wójcik, 1967; Kajstura et al., 2017; Koenig, 1990). For the purpose of analysis of the characteristics of the recorded test recordings, they were converted from AMR format into uncompressed WAVE PCM format with the help of Adobe Audition software. It was established that the frequency of sampling of all AMR recordings was 8000 Hz, and the frequency band of recorded sounds was within the range of approximately 200–3400 Hz. The occurrence of the so-called DC offset was not

ascertained in any of the test recordings. Some of the mobile phones used in the research had a pause function which could be used to pause the recording, which caused a discontinuity of the recording. On the basis of audio analysis and analysis of characteristics of AMR recordings, it was established that it is possible to reveal a discontinuity of the recording within the speech signal or other signals of known parameters. The above, however, is dependent on parameters of the mentioned signals, especially their signal-to-noise ratios. In recordings made with older telephone models, application of the pause function often caused recording of characteristic noises linked with the use of function keys. However, the use of touch screens in newer phones does not enable or significantly limits the recording of such noises. Characteristic traces in the structure of AMR files have not been ascertained after application of the pause function either. It has been observed that the so-called acoustic background of a recording in AMR format may be characterised by sudden changes in characteristics even though the recording is continuous and recorded in uniform acoustic conditions, which is dependent on the change in amplitude of recorded sounds. The AMR format was mainly designed for the transmission and recording of the speech signal, and the sound in the file is recorded only in the form of coefficients and is characterised by fairly strong compression. Accordingly, conclusions from an analysis of the integrity of an AMR recording using auditory analysis and analysis of characteristics of the recording should be verified by an additional method.

#### 4.2. Electrical network frequency analysis

One of the basic methods of investigating the authenticity of digital recordings is to analyse power line hum, changes in the frequency of which correspond to frequency fluctuations of the current in the electrical network (Grigoras, 2005). This is a method that

is widely recognised and applied by leading forensic audio laboratories; it can be described as a special type of analysis of the characteristics of a recording. Above all, it enables you to determine the continuity of a recording and the date and time that it was recorded (Grigoras et al., 2009; Kajstura et al., 2005; Michalek, 2009). Analyses of the characteristics of the test recordings in AMR format in the present research showed that in some of them it was possible to record the second harmonic of the power line hum, i.e., around 100 Hz frequency. Importantly, the amplitude and continuity of this signal enabled its extraction, calculation of instantaneous values of frequency and comparison with reference values. This harmonic was recorded in recordings made using various bit rates, and it was noted in recordings made on Nokia, Sony Ericsson and Samsung telephones. However, in none of the test recordings was the presence of a power line hum signal with a fundamental frequency, i.e., about 50 Hz, ascertained.

Fig. 1 shows the recorded second harmonic of the power line hum signal in an AMR recording made on a Sony Ericsson W595S telephone with a bit rate of 12.2 kb/s.

As part of the research, the influence of the bit rate of the AMR codec on the value of the amplitude of the second harmonic of the power line hum signal was also analysed. The amplitude of the power line hum is one of the most important parameters of this signal (Grigoras, 2005; Kajstura et al., 2005; Michalek, 2009). In order to do this, a test recording in uncompressed format, containing a harmonic of the power line hum with a frequency of about 100 Hz was converted with the help of the popular ffmpeg codec to recordings in AMR format using the available bit rates. The highest value of amplitude of the second harmonic was obtained for a recording coded with a bit rate of 12.2 kb/s, but a drop of  $-4$  dB was noted in relation to the harmonic amplitude from the uncompressed recording. For the bit rates: 10.2, 7.95, 7.4, 6.7, 5.9, 5.15

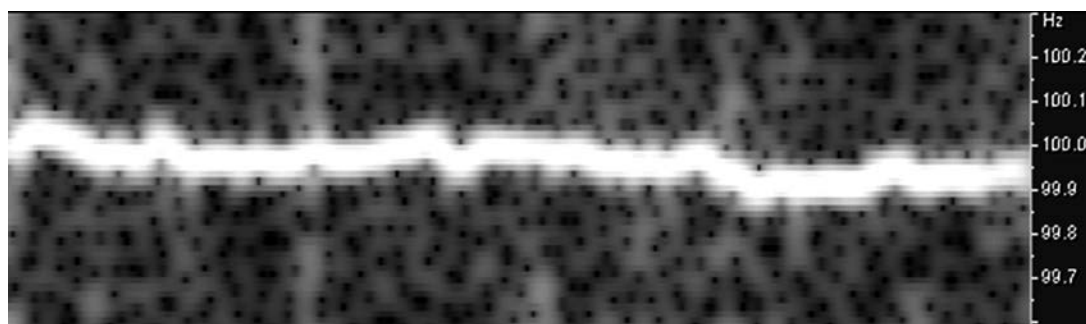


Fig. 1. Second harmonic of the power line hum signal in an AMR recording made using a Sony Ericsson W595S mobile phone.

and 4.75 kb/s, further drops in amplitude of this signal were noted, being, respectively:  $-4.35$ ,  $-4.65$ ,  $-6.75$ ,  $-7.7$ ,  $-8.9$ ,  $-9.8$  and  $-10$  dB in relation to the uncompressed recording. It follows that use of a lower bit rate during AMR coding, i.e., stronger compression, causes a reduction in the value of the amplitude of the second harmonic of the power line hum signal. It was found that a change in this amplitude did not have an effect on the correctness of verification of the time when the prepared recordings were made in relation to the reference database, on condition, however, of an acceptable signal-to-noise ratio (Michałek, 2009).

### 4.3. Analysis of AMR audio files properties

On the basis of the test recordings made in AMR format, it was established that each of them was saved to a file with the filename extension *\*.amr*.

The basic method of analysis of files, including audio files, is to read the information presented by the operating system of the computer. This includes: file type; timestamps for creation, modification and last access; and file size. For recognised formats, the operating system also gives parameters relating to the saving of the audio file such as: format, coding algorithm, bit rate, sampling frequency and quantization resolution (Hong, Yin, 2013; Kajstura et al., 2017; Michałek, 2016). Dedicated programmes are also available for reading detailed information about an audio file.

Using the File properties option of Windows operating systems – XP Professional, 7 Ultimate and 10 Professional – information was obtained about the AMR audio files recorded in this research. It was established that all the mentioned Windows operating systems only presented basic information on files in AMR format such as: names of files, their size, and type of file: AMR Narrow-Band Content, as well as the date and time of creation, modification and last access.

Several popular programmes designed for analysis of multimedia files were also used for analysis of the properties of files saved in AMR format: GSpot v. 2.70, MediaInfo v. 0.7.95 and ffprobe as part of ffmpeg v. N-81664-g6f062eb.

GSpot does not recognise files stored in AMR format and does not provide any detailed information about them. MediaInfo and ffprobe recognise the AMR format and present the following information correctly: file format (Adaptive Multi-Rate Narrowband), size, number of channels (1), sampling frequency (8000 Hz) and bit resolution (13 bits). Both programmes, i.e., MediaInfo and ffprobe, additionally provide the following information imprecisely: dura-

tion of recording and bit rate in kb/s. Table 1 presents a comparison of actual bit rate values of AMR test files in the range of 4.75 to 12.2 kb/s with values obtained using the mentioned programmes.

Table 1  
*Real values of AMR bit rates compared with values obtained using three popular software designated for multimedia files analysis*

AMR bit rate [kb/s]			
Frame header	GSpot	MediaInfo	ffprobe
4.75	file unknown	5.2	5
5.15	file unknown	5.6	5
5.9	file unknown	6.4	6
6.7	file unknown	7.2	7
7.4	file unknown	8.0	8
7.95	file unknown	8.4	8
10.2	file unknown	10.8	10
12.2	file unknown	12.8	12

### 4.4. Analysis of data structure in AMR format files

One of the methods applied during authenticity investigations of multi-media recordings is analysis of the data structure of files in which these recordings are saved. This analysis is carried out using programmes that enable visualisation and editing of files in so-called hexadecimal code. Multimedia files have metadata in their structure, which describe the data constituting the encoded recording. Metadata can contain significant information about a file, for example: the format of the recording and its parameters, the coding algorithm, timestamps and the name of the recording device. Analysis of the structure of the file also enables you to disclose any possible disturbances of its integrity, as well as entries indicating editing of the file with the help of computer software (Hong, Yin, 2013; Kajstura et al., 2017).

Analysis of the structure of files with test recordings in AMR format showed that each of them contains a 6 byte file header, which is saved as 23 21 41 4D 52 0A in hexadecimal code, which corresponds to `#!AMR\n` in ASCII code. An example AMR file header is presented in Fig. 2.

For the AMR-WB format, the file header is 23 21 41 4D 52 2D 57 42 0A in hexadecimal, which corresponds to `#!AMR-WB\n` in ASCII code.

Directly after the file header is the ordered frame structure. Each frame in an individual file contains a coded fragment of recording with a constant bit rate. At the beginning of each frame is its 1 byte header, which stores information about the coding mode, the bit rate and possible errors in the frame data. Audio in AMR format files is coded in eight modes with a bit rate that is dependent on the quality of recording: 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2 or 12.2 kb/s. Each frame contains a fragment of recording lasting 20 ms, but – depending on the AMR mode for the given device – the amount of bytes in each frame and its size vary. Table 2 presents information concerning possible AMR coding modes, bit rate, the size in bytes of frames containing coded sound, and the frame header in hexadecimal notation.

Table 2  
Values of AMR bit rates, frame sizes and frame headers, depending on AMR modes

AMR mode	AMR bit rate [kb/s]	Frame size [bytes]	Frame header [HEX]
AMR_4.75	4.75	13	04
AMR_5.15	5.15	14	0C
AMR_5.9	5.9	16	14
AMR_6.7	6.7	18	1C
AMR_7.4	7.4	20	24
AMR_7.95	7.95	21	2C
AMR_10.2	10.2	27	34
AMR_12.2	12.2	32	3C

Fig. 3 shows a portion of data from one of the AMR test files with the structure of the first five frames highlighted. As can be seen, each frame contains a header with a value of 14 in hexadecimal notation, which indicates that its size is 16 bytes, and the data are characterised by a bit rate of 5.9 kb/s.

Analysis of AMR files containing test recordings made using the voice recorder function showed that frame headers had values between 04 and 3C. Table 3 shows the obtained parameters of audio files recorded using the voice recorder function of the studied mobile phones.

It was established that all recordings originating from voice mailboxes of the cellular network operator PLAY which were recorded in AMR file format have a frame header with a value of 2C, and a bit rate of 7.95 kb/s. Table 4 presents the parameters of the mentioned files, including the brand and model of mobile phone on which they were saved.

As mentioned, the AMR file structure contains information about: the file format, coding mode, bit rate and possible errors in frames. However, analysis of this structure did not reveal the presence of data on: the phone brand and model, the date and time of recording, and the size of the file.

Due to the fact that in a single AMR file, each frame has the same header and length, they form an ordered data structure. On this basis, you can perform an analysis of the AMR file to assess its consistency and detect any possible disturbances or edits. Fig. 3 shows a fragment of data from one of the AMR test files that has a consistent structure. Fig. 4 shows the initial fragment of the same file, but – starting from the second frame – part of the data corresponding to about

```

00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F
23 21 41 4D 52 0A 14 A7 C9 B0 24 34 11 01 33 AC #!AMR..SÉ°$4..3~
D8 05 41 08 55 18 14 9B AC 65 87 A3 80 08 91 DB Ø.A.U..>~e+£€. 'Û
71 35 41 8D 61 40 14 95 87 50 AB 87 B1 03 68 E2 q5A.a@. *P«+±.há
    
```

Fig. 2. The file header of an AMR audio file.

```

00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F
23 21 41 4D 52 0A 14 A7 C9 B0 24 34 11 01 33 AC #!AMR..SÉ°$4..3~
D8 05 41 08 55 18 14 9B AC 65 87 A3 80 08 91 DB Ø.A.U..>~e+£€. 'Û
71 35 41 8D 61 40 14 95 87 50 AB 87 B1 03 68 E2 q5A.a@. *P«+±.há
B3 5E F8 41 4D 10 14 CE 1B B3 75 37 E2 A2 23 19 ²^øAM..Î.³u7âc#.
32 F1 F2 72 9A 8C 14 D7 4C BA B2 B7 90 36 47 62 2ñòršE.xL°². .6Gb
64 07 44 21 69 54 14 D7 D9 FF 9D EF 9A CB 9C C9 d.D!iT.×Ûÿ.išĚœĚ
2E D5 45 47 FB 38 14 EA 08 ED 24 2D F2 EC CA 59 .ÖEGú8.é.iš-òîËÿ
    
```

Fig. 3. Frames structure inside an AMR audio file.

Table 3  
AMR file parameters collected from tested devices:  
voice recorder function

Mobile phone: voice recorder function			
Brand and model	Frame header [HEX]	Frame size [bytes]	Bit rate [kb/s]
BlackBerry 8520	3C	32	12.2
BlackBerry 9320	3C	32	12.2
MaxCom MM461BB	3C	32	12.2
Nokia 3100c	14	16	5.9
Nokia 6103	14	16	5.9
Nokia 6230	3C	32	12.2
Nokia 6300	14	16	5.9
Nokia 6500c	3C	32	12.2
Nokia C1-01	3C	32	12.2
Nokia C2-02	3C	32	12.2
Nokia E50	0C	14	5.15
Nokia E52-1	3C	32	12.2
Nokia E65	0C	14	5.15
Nokia N95	0C	14	5.15
Nokia 2700	3C	32	12.2
Samsung Solid B2100	14	16	5.9
Samsung Galaxy ACE III	3C	32	12.2
Siemens C75	04	13	4.75
Sony Ericsson W595S	3C	32	12.2

0.2 seconds of recording has been removed from it. In the diagram, the second frame – from which audio data have been removed – is highlighted, and headers of frames in the processed file are boxed. As can be seen, as a result of this edition, the structure of the file has been disturbed.

On the basis of analysis of the structure of the test recordings, it was established that in AMR files recorded using a Nokia E52-1 telephone, there are disturbances in the data structure in correctly executed and

Table 4  
AMR file parameters collected from tested devices:  
voice mailboxes

Mobile phone: voice mailboxes (PLAY operator)			
Brand and model	Frame header [HEX]	Frame size [bytes]	Bit rate [kb/s]
HTC WildFire	2C	21	7.95
LG 9	2C	21	7.95
SAMSUNG GT-C3350	2C	21	7.95
SONY Ericsson WT19i	2C	21	7.95

continuous recordings. It was noticed that the length of some frames with coded sound did not correspond to the length indicated in their headers. There were various locations of these disturbances in successive files created on the mentioned mobile phone. Fig. 5 shows the initial fragment of an AMR file recorded with the above phone, which illustrates the above described problem. In this diagram, frame headers with a value of 3C are boxed and the incorrect seventh frame is highlighted. These types of disturbances in structure were not noted in AMR files recorded by the remaining studied phones.

The above described self-arising disturbances in AMR file structure do not mean that such a file has been edited using computer software. However, they indicate a characteristic feature for the analysed phone. That is why making test recordings with the studied recorder and performing a comparative analysis with evidential recordings are such significant elements of an authenticity examination (Michalek, 2016).

On the basis of analysis of recordings made for the purposes of this research paper, it was not ascertained that application of the pause function in the course of recording caused disturbances in data structure in AMR files.

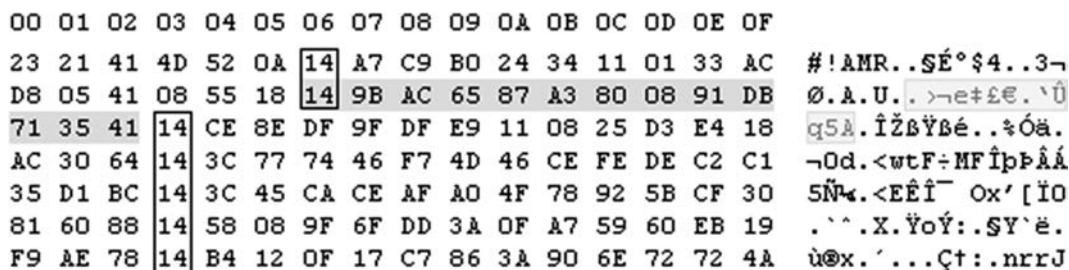


Fig. 4. Inconsistent structure of tampered AMR audio file.

### 5. Software for analysing AMR files

Based on the above described results of analysis of AMR file structure, our own algorithms and software were developed for automatic analysis of these types of audio files. This software was constructed and implemented in a MATLAB computational environment. It is presented in the form of a block diagram in Fig. 6.

The general principle of operation of the algorithm and software for automatic analysis of AMR files is as follows:

- 1) import of all the data saved in the audio file,
- 2) detection of file header and recognition of AMR type,
- 3) detection of headers of successive frames with coded sound, and determination of its parameters using the built-in database,
- 4) analysis of file structure frame after frame and comparison of parameters of all frames with each other on the basis of information read from headers,

- 5) calculation of the number of audio data and exact duration of recording on the basis of read parameters,
- 6) verification of file integrity and detection of possible disorders of structure or detection of frames with other parameters,
- 7) detailed report on the analysis of the file.

A significant element of the described software is the presentation of a report containing results of the automatic file analysis. This report contains detailed information about the parameters of the file and the recorded sound, the exact duration of the recording, the results of analysis of consistency of structure and the number of correct and any incorrect frames in the AMR file.

Fig. 7 shows a sample report generated on the basis of analysis of a consistent file named *Nagr002.amr* recorded with a Nokia 3100c mobile phone.

Fig. 8 shows the results of analysis of an AMR file recorded using a Nokia E52-1 phone, in which data structure disturbances occur.

00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F
23	21	41	4D	52	0A	3C	B3	25	0D	C8	BE	A1	81	88	08	E9	6F	E6	00	04	44	80	00	23	76	AE	94	00	00	00	03
4F	53	DF	E0	49	10	3C	20	84	5F	9F	D5	68	60	60	01	E2	8A	30	50	00	00	C0	00	01	20	00	01	3F	20	00	04
E0	60	00	10	0E	40	3C	91	13	D5	B6	66	79	E1	E0	01	E7	EA	F0	10	00	00	80	00	00	00	00	00	22	00	00	00
00	00	00	10	0C	80	3C	48	77	1D	8E	66	79	E1	E0	01	E7	BA	F0	00	00	00	C0	00	00	00	00	00	00	00	00	00
00	00	00	00	00	00	3C	55	00	83	B6	66	79	E1	E0	01	E7	CF	F0	40	00	00	80	00	01	14	00	00	00	00	00	00
28	D0	00	00	00	00	3C	E2	83	64	76	66	79	E1	E0	01	E7	BB	F0	10	00	01	40	00	00	00	00	00	30	80	00	00
00	00	00	03	90	00	3C	90	FB	21	D6	D1	79	E1	E0	01	E7	DF	F0	50	00	00	80	00	01	03	00	00	24	9C	00	00
4E	F0	00	10	20	00	44	22	A4	5C	B0	0E	7C	7C	44	22	A5	7C	B0	1E	3C	7D	0D	5D	D0	18	43	81	EC	A2	CF	1F
F7	74	23	15	80	00	5F	6F	FA	33	A6	58	00	05	CA	D9	7C	D0	4D	20	3C	68	77	1E	A2	A5	CC	09	65	AA	29	D5
19	F1	2B	0E	A9	13	8E	E2	20	08	F0	F8	DD	33	BD	76	7A	55	B1	C0	3C	3B	03	5D	B7	E7	50	12	C1	EF	2A	CA
0E	B9	28	37	5A	87	59	4A	70	CF	38	44	82	8A	A5	0F	3A	EF	28	30	3C	44	72	80	BA	FF	7C	02	81	FA	4C	49
B7	C0	AC	33	C3	2E	CA	87	06	D5	31	3E	2E	A0	5B	C9	56	7D	65	90	3C	70	81	60	E6	A3	D2	21	C3	DC	A3	F8

Fig. 5. Disturbances in the structure of an AMR file recorded by a Nokia E52-1 mobile phone.

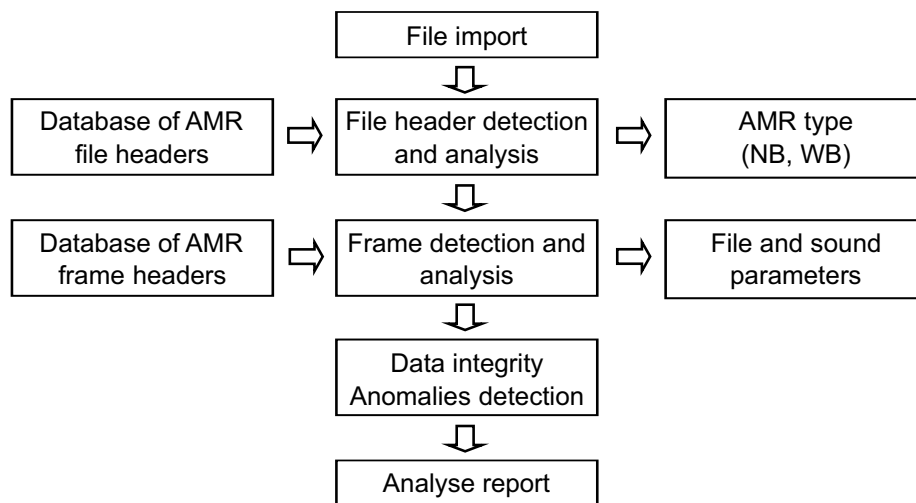


Fig. 6. A block scheme of the software constructed for automatic analysis of data structure of AMR audio files.



## Conclusions

Performing research on the properties of audio recorded in AMR format enabled us to acquire and systematize knowledge on these types of recordings and files, and assess whether such a format is suitable for authentication investigations. It turned out that recording in the seemingly uncomplicated AMR format enabled authentication examinations; the results of the performed research indicated that for such recordings, it is possible to apply generally recognised methods of authentication analysis, such as – primarily – analysis of the power line hum signal and file data structure. Auxiliary methods may include analysis of characteristics of the recording and auditory analysis. The Windows operating system and popular programmes

for analysis of multimedia files only provide basic and often imprecise information about files saved in AMR format. Based on results of performed analyses of AMR file structures, our own software was developed and implemented for automatic analysis of these types of files working in the MATLAB computational environment. This software determines the correct parameters and analyses the integrity of files, as well as detecting any possible structural disturbances. The collected and analysed test recordings, together with the developed software for automatic analysis of AMR files constitute useful tools for authentication examinations of evidential recordings, which are carried out as part of expert witnessing at the Section of Speech and Audio Analysis at the Institute of Forensic Research.

```

Command Window
File name: Nagr002.amr
File type: AMR-NB
Audio parameters: Fs=8000Hz, nbits=13, 1 channel
File header: #!AMR\n (correct)
Frame header (hex): 14
Frame Type and Mode Indication: 2
Frame content and bandwidth: AMR 5,90 kbit/s
Frame length [bytes]: 16(dec), 10(hex)
Predicted number of frames: 160
Number of correct frames: 160
Number of incorrect frames: 0
! AMR file is consistent !
Total time of the recording is [s]: 3.2
fx >>

```

Fig. 7. Detailed report of the analysis of a consistent AMR audio file recorded using a Nokia 3100c mobile phone.

```

Command Window
File name: Plik audio 05.amr
File type: AMR-NB
Audio parameters: Fs=8000Hz, nbits=13, 1 channel
File header: #!AMR\n (correct)
Frame header (hex): 3C
Frame Type and Mode Indication: 7
Frame content and bandwidth: AMR 12,2 kbit/s
Frame length [bytes]: 32(dec), 20(hex)
Predicted number of frames: 2101
Number of correct frames: 1975
Number of incorrect frames: 41
! WARNING: AMR file is NOT consistent !
Total time based on correct frames is [s]: 39.5
fx >>

```

Fig. 8. Detailed report of the analysis of an inconsistent AMR audio file recorded using a Nokia E52-1 mobile phone.

## References

1. 3rd Generation Partnership Project (3GPP). *Technical Specification Group Services and System Aspects*. (2010). Technical Specification 26.244 V9.2.0: Transparent end-to-end packet switched streaming service (PSS). <http://www.3gpp.org> (access 05.07.2017).
2. 3rd Generation Partnership Project (3GPP). *Technical Specification Group Services and System Aspects* (2011). Technical Specification 3GPP TS 26.071 V10.0.0: Mandatory speech CODEC speech processing functions; AMR speech CODEC; General description. <http://www.3gpp.org> (accessed 05.07.2017).
3. 3rd Generation Partnership Project (3GPP). *Technical Specification Group Services and System Aspects*. (2011). Technical Specification 3GPP TS 26.090 V10.1.0: Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Transcoding functions. <http://www.3gpp.org> (accessed 05.07.2017).
4. Bhatt, N. S., Kosta, Y. P. (2011). Architectural study, implementation and objective evaluation of code excited linear prediction based GSM AMR 06.90 speech coder using MATLAB. *International Journal of Advanced Engineering Technology*, 2(2), 52–59.
5. Błasikiewicz, S., Miściuk, A., Wójcik, W. (1967). Podstawowy zakres badań fonoskopijnych prowadzonych w Zakładzie Kryminalistyki KG MO. *Problemy Kryminalistyki*, 67/68, 303–327.
6. European Telecommunications Standards Institute (ETSI). (2002). Technical Specification ETSI TS 126 071 V5.0.0: Universal Mobile Telecommunications System (UMTS); AMR speech Codec; General description. <http://www.etsi.org> (accessed 01.07.2017).
7. Gloe, T., Fischer, A., Kirchner, M. (2014). Forensic analysis of video file formats. *Digital Investigation*, 11, 68–76.
8. Grigoras, C. (2005). Digital audio recording analysis: The electric network frequency (ENF) criterion. *International Journal of Speech, Language and the Law*, 12, 64–76.
9. Grigoras, C., Cooper, A., Michalek, 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, ENFSI Forensic Speech and Analysing Working Group, [www.enfsi.eu](http://www.enfsi.eu) (accessed 05.07.2017).
10. Hong, G., Yin, Z. (2013). Research on digital audio authenticity analysis. (In) *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering* (ICCSEE 2013). Atlantis Press.
11. Kajstura, M., Michalek, M., Trawińska, A. (2017). Ekspertyza fonoskopijna. (In) M. Kała, D. Wilk, J. Wójcikiewicz (Eds.), *Ekspertyza Sądowa. Zagadnienia wybrane*, (674–726). Warszawa: Wolters Kluwer.
12. Kajstura, M., Trawińska, A., Hebenstreit, J. (2005). Application of the electrical network frequency (ENF) criterion; A case of a digital recording. *Forensic Science International*, 155, 165–171.
13. Koenig, B. E. (1990). Authentication of forensic audio recordings. *Journal of the Audio Engineering Society*, 38(1/2), 3–33.
14. Luo, D., Yang, R., Huang, J. (2015). Identification of AMR decompressed audio. *Digital Signal Processing*, 37, 85–91.
15. Michalek, M. (2009). The application of powerline hum in digital recording authenticity analysis. *Problems of Forensic Sciences*, 80, 355–364.
16. Michalek, 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.
17. Sjöberg, J., Westerlund, M., Lakaniemi, A., Xie, Q. (2007). *Request for Comments 4867: RTP Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs*. Network Working Group, The IETF Trust, [www.ietf.org](http://www.ietf.org) (accessed 11.07.2017).

## Corresponding author

Dr inż. Marcin Michalek  
 Section of Speech and Audio Analysis  
 Institute of Forensic Research  
 ul. Westerplatte 9  
 PL 31-033 Kraków  
 e-mail: mmichalek@ies.krakow.pl

# WŁAŚCIWOŚCI NAGRAŃ I PLIKÓW DŹWIĘKOWYCH ZAPISANYCH W FORMACIE AMR ORAZ OCENA MOŻLIWOŚCI ICH ZASTOSOWANIA W BADANIACH AUTENTYCZNOŚCI

## 1. Wstęp

Kodek *Adaptive Multi-Rate* (AMR) został ustandaryzowany przez *European Telecommunications Standards Institute* (ETSI) i jest rozwijany od 1999 roku przez *3rd Generation Partnership Project* (3GPP). Występuje on w dwóch odmianach, tj. wąskopasmowy AMR (*Narrowband*), określane również jako AMR-NB, oraz szerokopasmowy AMR-WB (*Wideband*, 3GPP, 2011a; ETSI, 2002). AMR i AMR-WB były pierwotnie przeznaczone do zastosowania w systemach komunikacji radiowej. Z uwagi na elastyczność tego kodeka i możliwość jego adaptacji do różnych jakości połączeń komunikacyjnych znalazł on także zastosowanie w mobilnej telefonii komórkowej GSM i UMTS (3GPP, 2011b; Sjöberg, Westerlund, Lakaniemi, Xie, 2007). Oprócz zastosowania w systemach komunikacji mobilnej AMR jest wykorzystywany także do przesyłania mowy w audio- i wideokonferencjach oraz w łączności satelitarnej i VoIP. Zapewnia on akceptowalną jakość sygnału mowy nawet dla słabych parametrów połączenia (Bhatt, Kosta, 2011; Sjöberg i in., 2007). Format ten znajduje także zastosowanie do odtwarzania mowy w audiobookach oraz w komunikatach głosowych.

AMR to także format zapisu nagrań dźwiękowych, który wykorzystuje wymieniony wyżej kodek w wersji wąskopasmowej (3GPP, 2011a; Sjöberg i in., 2007). Nagrania dźwiękowe w tym formacie rejestrowane są głównie za pomocą telefonów komórkowych, ale także przez inne popularne urządzenia mobilne, na przykład tablety (Luo, Yang, Huang, 2015; Michałek, 2016). AMR umożliwia zapis nagrań do plików dźwiękowych o różnej jakości z wykorzystaniem przepływności od 4,75 do 12,2 kb/s (3GPP, 2011a). Format ten może być także stosowany do utrwalania dźwięku wewnątrz struktury tak zwanych kontenerów multimedialnych, w których zapisywane są nagrania wideo lub dźwiękowe. Przykładem są pliki w formacie 3GP i MP4, bardzo popularne w nowoczesnych urządzeniach mobilnych, a szczególnie w telefonach komórkowych (3GPP, 2010; Gloe, Fischer, Kirchner, 2014).

W niniejszej pracy skoncentrowano się na właściwościach nagrań i plików dźwiękowych utrwalanych w wąskopasmowym formacie AMR oraz możliwości ich wykorzystania w analizie autentyczności nagrań. Badania wykonano z wykorzystaniem nagrań dźwiękowych zarejestrowanych za pomocą telefonów komórkowych

z funkcją dyktafonu oraz zapisanych w tzw. skrzynkach głosowych operatorów sieci komórkowych.

Aktualnie badania autentyczności nagrań cyfrowych wymagają konkatencji wyników i wniosków uzyskanych z różnych fonoskopijnych metod badawczych, często przy współdziałaniu metod z zakresu informatyki sądowej (Kajstura, Michałek, Trawińska, 2017). Wyniki wykonanych prac wskazują, że rejestracja nagrań w stosunkowo nieskomplikowanym formacie AMR pozwala na wykorzystanie uznanych i stosowanych metod do weryfikacji autentyczności, takich jak analiza przydźwięku sieciowego oraz właściwości i struktury pliku (Hong, Yin, 2013; Grigoras, 2005; Kajstura i in., 2017).

Badania opisane w niniejszym artykule wykonano w ramach projektu badawczego zrealizowanego w Instytucie Ekspertyz Sądowych im. Prof. dra Jana Sehna w Krakowie.

## 2. Kodowanie dźwięku w plikach formatu AMR

Pliki formatu AMR nie zawierają zakodowanego dźwięku próbka po próbce, ale w postaci współczynników opisujących kolejne fragmenty nagrania. Kodek ten wykorzystuje algorytm liniowej predykcji zwany *Algebraic Codebook Excitation Linear Prediction* (ACELP), który jest modyfikacją popularnego algorytmu liniowego kodowania predykcyjnego *Linear Predictive Coding* (LPC) stosowanego do kodowania i opisu sygnału mowy (3GPP, 2011b; Luo i in., 2015).

Przed właściwym procesem kodowania dźwięku za pomocą AMR następuje jego wstępne przygotowanie, które składa się z konwersji analogowego sygnału do postaci cyfrowej o parametrach: rozdzielczość 13 bitów, częstotliwość próbkowania 8000 Hz i kodowanie *Pulse Code Modulation* (PCM). Sygnał poddawany jest także filtracji górno-przepustowej z częstotliwością odcięcia wynoszącą 80 Hz (3GPP, 2011b).

Kodek AMR działa w ten sposób, że dokonuje podziału całego zapisywanego nagrania na krótsze fragmenty, tzw. ramki, o czasie trwania 20 ms. Dla częstotliwości próbkowania 8000 Hz pojedyncza ramka zawiera 160 próbek. W każdej ramce analizowany jest sygnał mowy i wyznaczone są parametry modelu ACELP: współczynniki LPC, adaptacyjne i stałe książki kodowe oraz wzmocnienia. Aby odtworzyć treść zapisanego nagrania, dekodery na podstawie parametrów modelu ACELP wy-

konuje proces syntezy zakodowanego sygnału (3GPP, 2011a; ETSI, 2002).

W dalszej części przedstawiono pokrótce zasadę działania kodera AMR na podstawie (3GPP, 2011a i 2011b; ETSI, 2002; Sjoberg i in., 2007). Proces zapisu dźwięku w pliku z wykorzystaniem kodeka AMR można podzielić na kilka głównych etapów. W pierwszym z nich wykonywana jest analiza krótkookresowa i obliczane są współczynniki liniowej predykcji LPC 10. rzędu za pomocą algorytmu Levinsona-Durbina zgodnie ze wzorem:

$$H(z) = \frac{1}{1 + \sum_{i=1}^m \hat{a}_i z^{-i}} \quad (1)$$

gdzie:  $\hat{a}_i$  to współczynniki predykcji liniowej, a  $m=10$  to rząd predykcji.

Zależność (1) umożliwia krótkookresową analizę sygnału, w której modelowany jest trakt wokalny. W etapie tym wykonywana jest także analiza długookresowa i obliczane są parametry funkcji pobudzenia traktu wokalnego w celu późniejszej syntezy sygnału mowy za pomocą zależności:

$$\frac{1}{B(z)} = \frac{1}{1 - g_p z^{-T}} \quad (2)$$

gdzie:  $T$  to okres pobudzenia, a  $g$  to jego wzmocnienie.

Na podstawie analizy długookresowej tworzona jest tak zwana adaptacyjna książka kodowa (*Adaptive Codebook*), którą dokładniej opisano poniżej.

W kolejnym etapie kodowania następuje transformacja obliczonych już współczynników LPC do postaci par LSP (*Line Spectral Pair*) na potrzeby dalszej kwantyzacji. Transformacja ta wykonywana jest na podstawie dekompozycji funkcji  $A(z)$  do wielomianów  $F_1'$  i  $F_2'$  według zależności:

$$\begin{aligned} F_1'(z) &= A(z) + z^{-11}A(z^{-1}) \\ i \ F_2'(z) &= A(z) - z^{-11}A(z^{-1}) \end{aligned} \quad (3)$$

Następnie tak wyznaczone współczynniki LSP zostają poddane kwantyzacji w dziedzinie częstotliwości zgodnie z zależnością:

$$f_i = \frac{f_s}{2\pi} \arccos(q_i) \quad (4)$$

gdzie:  $i=1, \dots, 10$ ,  $f_i$  to częstotliwości LSF (*Line Spectral Frequencies*) w zakresie  $[0, 4000]$  Hz odpowiadające współczynnikom LSP,  $f_s$  to częstotliwość próbkowania 8000 Hz, zaś  $q_i$  to współczynniki LSP w domenie kosinusowej. Następnie tak skwantyzowane współczynniki zostają poddane liniowej interpolacji, po czym z powrotem przetransformowane jako skwantyzowane współczynniki LPC.

W kolejnym etapie pojedyncza ramka zawierająca 160 próbek zostaje podzielona na cztery równe części, tzw. podramki (*subframes*), po 40 próbek każda i trwające po 5 ms. W zależności od przepływności bitowej kodeka AMR wykorzystuje się skwantyzowane i nieskwantyzowane współczynniki LPC oraz interpolowane współczynniki LSP w różnych podramkach. Na podstawie podramek wyznaczane są parametry ksiąg kodowych *Adaptive Codebook* oraz *Fixed Codebook*. *Adaptive Codebook* jest wyznaczana na podstawie analizy długookresowej i zawiera wektory pobudzeń o zmiennych wartościach w celu dostosowywania do parametrów sygnału mowy dla każdej podramki. Natomiast *Fixed Codebook* zawiera wektory pobudzeń o stałych wartościach. W końcowym etapie kodowania AMR wykorzystuje się techniki wyznaczania pobudzeń i ich wzmocnień oraz interwałów między pobudzeniami dla analizowanych ramek (*Open-loop Pitch Search* i *Closed-loop Pitch Analysis*) w celu wyznaczenia najbardziej optymalnych parametrów kodowanego sygnału mowy.

Zapis nagrania do pliku odbywa się w taki sposób, że dla każdej ramki o czasie trwania 20 ms koder generuje 95, 103, 118, 134, 148, 159, 204 albo 244 bity dla przepływności 4,75; 5,15; 5,9; 6,7; 7,4; 7,95; 10,2 albo 12,2 kb/s. Po zakodowaniu pojedynczej ramki pamięć kodera jest aktualizowana w celu obliczenia parametrów dla następnej ramki. Struktura każdego pliku AMR składa się z jego nagłówka, po którym znajduje się uporządkowany zbiór ramek zawierający zakodowane kolejne fragmenty nagrania (3GPP, 2011b; Sjoberg i in., 2007).

### 3. Materiał do badań i zastosowane metody

W celu wykonania zamierzonych badań zgromadzone i przeanalizowano możliwości rejestracji dwudziestu pięciu telefonów komórkowych następujących marek i modeli:

- BlackBerry: 8520 i 9320;
- HTC Wildfire;
- LG 9;
- MaxCom MM461BB;
- Nokia: 2700, 3100c, 6103, 6230, 6300, 6500c, C1-01, C2-02, E50, E52-1, E65 i N95;
- Samsung: Galaxy Ace III, GT-C3350 i Solid B2100;
- Siemens C75;
- Sony Ericsson: W595S i WT19i.

Za pomocą wymienionych wyżej urządzeń wykonano 181 nagrań testowych zapisanych w formacie AMR. Wykorzystano do tego celu funkcję dyktafonu oraz rejestrację komunikatów w skrzynkach głosowych operatorów sieci komórkowych, które następnie zapisano do pamięci telefonów w plikach formatu AMR. Nagrania te wykonano z wykorzystaniem dostępnych funkcji podczas rejestracji, tj. uruchomienia i zakończenia nagry-

wania oraz funkcji pauzy. Zrealizowane one zostały w różnych warunkach akustycznych, tj w zamkniętych pomieszczeniach oraz na otwartej przestrzeni. Zawierają zakres dźwięków oraz celowych zakłóceń o możliwie szerokim spektrum częstotliwościowym. Po utrwaleniu w pamięci telefonu każde z nagrań testowych zostało skopiowane na dysk twardy komputera laboratoryjnego lub wykonano obraz zawartości pamięci urządzenia.

W celu dokonania oceny możliwości wykorzystania parametrów nagrań i plików utrwalonych w formacie AMR w badaniach autentyczności nagrania testowe poddano następującym analizom:

- charakterystyk czasowych i częstotliwościowych z jednoczesną analizą audytywną;
- sygnału przydźwięku sieciowego – jako szczególnie przypadek analizy charakterystyk nagrania;
- właściwości plików odczytywanych za pomocą systemu operacyjnego Windows oraz narzędzi przeznaczonych do analizy plików multimedialnych;
- struktury danych zapisanych w plikach formatu AMR z wykorzystaniem programów do wizualizacji i edycji w postaci kodu szesnastkowego i ASCII.

## 4. Wyniki

### 4.1. Analiza charakterystyk i audytywna

Analiza charakterystyk czasowych i częstotliwościowych połączona z analizą audytywną jest metodą badawczą stosowaną głównie do oceny ciągłości nagrania (Błasiakiewicz, Miściuk, Wójcik, 1967; Kajstura i in., 2017; Koenig, 1990). W celu analizy charakterystyk zarejestrowanych nagrań testowych przekonwertowano je z formatu AMR do nieskompresowanego formatu WAVE PCM za pomocą oprogramowania Adobe Audition. Ustalono, że częstotliwość próbkowania wszystkich nagrań AMR wynosi 8000 Hz, a pasmo częstotliwościowe utrwalonych dźwięków zawiera się w granicach, w przybliżeniu, 200–3400 Hz. W żadnym z nagrań testowych nie stwierdzono występowania tzw. składowej stałej sygnału. Część wykorzystanych do badań telefonów komórkowych umożliwia wstrzymanie nagrywania za pomocą funkcji pauzy, co powodowało brak ciągłości zapisu nagrania. Na podstawie analizy audytywniej i charakterystyk nagrań AMR ustalono, że ujawnienie braku ciągłości możliwe jest w obrębie sygnału mowy lub innych sygnałów o znanych parametrach. Jest to jednak zależne od parametrów wymienionych sygnałów, zwłaszcza stosunku ich amplitudy do zakłóceń. W nagraniach wykonanych za pomocą starszych modeli telefonów zastosowanie pauzy powodowało często rejestrację charakterystycznych odgłosów związanych z użyciem przycisków funkcyjnych. Wykorzystanie w nowszych telefonach ekranów dotykowych uniemożliwia lub znacznie ogranicza rejestrację takich

odgłosów. Nie stwierdzono też żadnych charakterystycznych śladów w strukturze plików AMR po zastosowaniu funkcji pauzy. Zaobserwowano, że tzw. tło akustyczne w nagraniu formatu AMR może odznaczać się nagłymi zmianami charakterystyk pomimo tego, że nagranie jest ciągłe i utrwalone w jednolitych warunkach akustycznych, co zależne jest od zmiany amplitudy rejestrowanych dźwięków. Format AMR zaprojektowany był głównie do przesyłania i rejestracji sygnału mowy, a dźwięk w pliku zapisany jest jedynie w postaci współczynników i odznacza się dość silną kompresją. W związku z powyższym wnioski z badania integralności nagrania AMR za pomocą analizy audytywniej i jego charakterystyk powinny być zweryfikowane dodatkową metodą.

### 4.2. Analiza sygnału przydźwięku sieciowego

Jedną z podstawowych metod badania autentyczności nagrań cyfrowych jest analiza sygnału przydźwięku sieciowego, którego zmiany częstotliwości odpowiadają wahaniom częstotliwości prądu w sieci elektroenergetycznej (Grigoras, 2005). Jest to metoda powszechnie uznana i stosowana przez wiodące laboratoria fonoskopijne; można stwierdzić, że jest ona szczególnie przypadkiem analizy charakterystyk nagrania. Umożliwia przede wszystkim ustalenie ciągłości zapisu nagrania oraz daty i godziny jego rejestracji (Grigoras, Cooper, Michałek, 2009; Kajstura, Trawińska, Hebenstreit, 2005; Michałek, 2009). Analiza charakterystyk nagrań testowych w formacie AMR wykazała, że w niektórych z nich możliwe było zarejestrowanie drugiej harmonicznej sygnału przydźwięku sieciowego, tj. wokół częstotliwości 100 Hz. Co ważne, amplituda i ciągłość tego sygnału pozwalała na jego ekstrakcję, obliczenie chwilowych wartości częstotliwości i porównanie z wartościami referencyjnymi. Drugą harmoniczną zarejestrowano w nagraniach z różnymi przepływnościami bitowymi i odnotowano ją w nagraniach wykonanych telefonami Nokia, Sony Ericsson oraz Samsung. W żadnym z nagrań testowych nie stwierdzono natomiast obecności sygnału przydźwięku sieciowego o częstotliwości podstawowej, tzn. wokół 50 Hz.

Na rysunku 1 przedstawiono zarejestrowaną drugą harmoniczną sygnału przydźwięku sieciowego w nagraniu AMR zarejestrowanym telefonem Sony Ericsson W595S z przepływnością 12,2 kb/s.

W ramach prac wykonano także analizę wpływu przepływności bitowej kodeka AMR na wartość amplitudy drugiej harmonicznej sygnału przydźwięku sieciowego. Amplituda przydźwięku to jeden z najważniejszych parametrów tego sygnału (Grigoras, 2005; Kajstura i in., 2005; Michałek, 2009). W tym celu nagranie testowe w nieskompresowanym formacie, zawierające harmoniczną przydźwięku wokół częstotliwości 100 Hz, przekonwertowano za pomocą popularnego kodeka f1mpeeg

do nagrań w formacie AMR z wykorzystaniem możliwych przepływności. Największą wartość amplitudy drugiej harmonicznej uzyskano dla nagrania zakodowanego z przepływnością 12,2 kb/s, przy czym zauważono jej spadek o  $-4$  dB w stosunku do amplitudy harmonicznej z nagrania nieskompresowanego. Dla przepływności: 10,2; 7,95; 7,4; 6,7; 5,9; 5,15 i 4,75 kb/s zanotowano dalsze spadki amplitudy tego sygnału wynoszące odpowiednio  $-4,35$ ;  $-4,65$ ;  $-6,75$ ;  $-7,7$ ;  $-8,9$ ;  $-9,8$  i  $-10$  dB w odniesieniu do nagrania nieskompresowanego. Wynika z tego, że zastosowanie mniejszej przepływności bitowej podczas kodowania AMR, czyli silniejszej kompresji, powoduje redukcję wartości amplitudy drugiej harmonicznej sygnału przydźwięku sieciowego. Ustalono, że zmiana tej amplitudy nie miała wpływu na poprawność weryfikacji czasu rejestracji przygotowanych nagrań względem bazy referencyjnej, pod warunkiem jednak akceptowalnego stosunku amplitudy sygnału do zakłóceń (Michalek, 2009).

#### 4.3. Analiza właściwości plików dźwiękowych AMR

Na podstawie wykonanych nagrań testowych w formacie AMR ustalono, że każde z nich zapisane zostało do pliku z rozszerzeniem *\*.amr*.

Podstawową metodą analizy plików, w tym także dźwiękowych, jest odczytanie informacji prezentowanych przez system operacyjny komputera. Są to przede wszystkim: typ pliku, sygnatury czasowe utworzenia, modyfikacji i ostatniego dostępu oraz jego rozmiar. Dla rozpoznawanych formatów system operacyjny podaje również parametry zapisu pliku dźwiękowego, takie jak: format, algorytm kodowania, przepływność, częstotliwość próbkowania i rozdzielczość kwantowania (Hong, Yin, 2013; Kajstura i in., 2017; Michalek, 2016). Do odczytu szczegółowych informacji pliku dźwiękowego służą również dedykowane programy.

Za pomocą opcji *Właściwości pliku* systemów operacyjnych Windows: XP Professional, 7 Ultimate oraz 10 Professional odczytano informacje dotyczące plików dźwiękowych AMR zarejestrowanych w ramach niniejszej pracy. Ustalono, że wszystkie wymienione systemy operacyjne Windows dla plików w formacie AMR prezentowały jedynie podstawowe informacje, takie jak: nazwy plików, ich rozmiar, typ pliku *AMR Narrow-Band Content* oraz daty i godziny utworzenia, modyfikacji i ostatniego dostępu.

Do analizy właściwości plików zapisanych w formacie AMR wykorzystano także kilka popularnych programów przeznaczonych do analizy plików multimedialnych, tj.: GSpot v. 2.70, MediaInfo v. 0.7.95 oraz fprobe będący częścią pakietu ffmpeg v. N-81664-g6f062eb.

Program GSpot nie rozpoznaje plików zapisanych w formacie AMR i nie podaje o nich żadnych szczegółowych informacji. Programy MediaInfo oraz fprobe rozpoznają format AMR i prezentują poprawnie następujące informacje: format pliku (*Adaptive Multi-Rate Narrowband*), rozmiar, liczbę kanałów (1), częstotliwość próbkowania (8000 Hz) oraz rozdzielczość bitową (13 bitów). Oba programy, tj. MediaInfo i fprobe, podają dodatkowo nieprecyzyjnie: czas trwania nagrania oraz wartość przepływności bitowej w kb/s. W tabeli 1 przedstawiono porównanie rzeczywistych wartości przepływności bitowej testowych plików AMR z zakresu od 4,75 do 12,2 kb/s z wartościami odczytanymi za pomocą wymienionych programów.

#### 4.4. Analiza struktury danych w plikach formatu AMR

Jedną z metod stosowanych podczas badań autentyczności nagrań multimedialnych jest analiza struktury danych w plikach z zapisem tych nagrań. Analizę taką przeprowadza się z wykorzystaniem oprogramowań umożliwiających wizualizację i edycję plików w tak zwanym kodzie szesnastkowym. Pliki multimedialne mają w swojej strukturze metadane, które opisują właściwe dane z zakodowanym nagraniem. Metadane mogą zawierać istotne informacje o pliku, przykładowo: format zapisu i jego parametry, algorytm kodowania, sygnatury czasowe lub nazwę urządzenia rejestrującego. Analiza struktury pliku umożliwi również ujawnienie ewentualnego zaburzenia jego integralności oraz wpisów wskazujących na edycję pliku za pomocą oprogramowania (Hong, Yin, 2013; Kajstura i in., 2017).

Analiza struktury plików z nagraniami testowymi w formacie AMR wykazała, że każdy z nich składa się z nagłówka pliku o rozmiarze 6 bajtów, który zapisany jest jako 23 21 41 4D 52 0A w kodzie szesnastkowym, co odpowiada *#!AMR*\n w kodzie ASCII. Nagłówek przykładowego pliku AMR zaprezentowano na rysunku 2.

Dla formatu AMR-WB nagłówek pliku ma postać 23 21 41 4D 52 2D 57 42 0A w kodzie szesnastkowym, co odpowiada *#!AMR-WB*\n w kodzie ASCII.

Bezpośrednio za nagłówkiem pliku znajduje się uporządkowana struktura ramek. Każda ramka w pojedynczym pliku zawiera zakodowany fragment nagrania ze stałą przepływnością bitową. Na początku każdej ramki znajduje się jej nagłówek o rozmiarze 1 bajta, w którym zapisane są informacje o trybie kodowania, przepływności bitowej i ewentualnych błędach w danych ramki. Dźwięk w plikach formatu AMR kodowany jest w ośmiu trybach (zwanym również *modami*) z przepływnościami bitowymi zależnymi od jakości nagrywania: 4,75; 5,15; 5,9; 6,7; 7,4; 7,95; 10,2 albo 12,2 kb/s. Każda ramka zawiera fragment nagrania trwający 20 ms, jednak w zależności od trybu AMR dla danego urządzenia liczba baj-

tów w każdej ramce oraz jej rozmiar są różne. W tabeli 2 podano zestawienie informacji dotyczących możliwych trybów kodowania AMR, przepływności bitowej, rozmiaru ramki w bajtach z zakodowanym dźwiękiem oraz nagłówek ramki w zapisie szesnastkowym.

Na rysunku 3 przedstawiono fragment danych jednego z testowych plików AMR z zaznaczoną strukturą pięciu pierwszych ramek. Jak można zauważyć, każda ramka zawiera nagłówek o wartości 14 w zapisie szesnastkowym, co wskazuje, że jej rozmiar wynosi 16 bajtów, a dane charakteryzują się przepływnością 5,9 kb/s.

Analiza plików AMR z nagraniami testowymi zarejestrowanymi funkcją dyktafonu wykazała, że nagłówki ramek przyjmowały wartości od 04 do 3C. W tabeli 3 zaprezentowano odczytane parametry plików dźwiękowych utwalonych za pomocą funkcji dyktafonu dla przeanalizowanych telefonów komórkowych.

Ustalono, że wszystkie nagrania pochodzące ze skrzynek głosowych operatora sieci komórkowej PLAY, które zapisano w plikach formatu AMR, odznaczają się nagłówkiem ramek o wartości 2C i przepływnością bitową 7,95 kb/s. W tabeli 4 przedstawiono parametry wymienionych plików z uwzględnieniem telefonów komórkowych, w pamięci których zostały one zapisane.

Jak wspomniano, w strukturze plików AMR znajdują się informacje dotyczące: formatu pliku, trybu kodowania, przepływności bitowej i ewentualnych błędów w ramkach. Analiza tej struktury nie wykazała natomiast obecności danych związanych z marką i modelem urządzenia, datą i godziną rejestracji nagrania oraz rozmiarem pliku.

Ponieważ w pojedynczym pliku AMR każda ramka ma taki sam nagłówek oraz długość, tworzą one uporządkowaną strukturę danych. Na tej podstawie można wykonywać analizę pliku AMR w celu oceny jego spójności oraz detekcji ewentualnych zaburzeń lub edycji. Na rysunku 3 przedstawiono fragment danych jednego z testowych plików AMR mającego spójną strukturę. Na rysunku 4 zaprezentowano początkowy fragment tego samego pliku, przy czym licząc od drugiej ramki usunięto z niego część danych odpowiadającą około 0,2 sekundy nagrania. Na rysunku zaznaczono drugą ramkę, z której usunięto dane dźwiękowe, jak również nagłówki ramek w tak przetworzonym pliku. Jak można zauważyć, w wyniku tej edycji struktura pliku uległa zaburzeniu.

Na podstawie analizy struktury nagrań testowych ustalono, że w plikach AMR zarejestrowanych za pomocą telefonu Nokia E52-1 znajdują się zaburzenia struktury danych w prawidłowo wykonanych i ciągłych nagraniach. Zauważono, że długość niektórych ramek z zakodowanym dźwiękiem nie odpowiada długości wskazanej w ich nagłówkach. Miejsca występowania tych zaburzeń były różne w kolejnych plikach wykonanych tym samym telefonem komórkowym. Na rysunku 5 przedstawiono początkowy fragment pliku AMR zarejestrowanego tym

urządzeniem, który obrazuje opisany wyżej problem. Na niniejszym rysunku zaznaczono nagłówki ramek o wartości 3C oraz niepoprawną siódmą ramkę. Nie odnotowano tego typu zaburzeń struktury w plikach AMR zapisanych przez pozostałe badane urządzenia.

Opisane wyżej samoistne zaburzenia struktury pliku AMR nie oznaczają, że plik taki został poddany edycji za pomocą oprogramowania komputerowego. Wskazują natomiast na cechę charakterystyczną dla analizowanego urządzenia. Dlatego tak istotnym elementem w trakcie badań autentyczności jest wykonanie nagrań testowych badanym rejestratorem i ich analiza porównawcza z nagraniami dowodowymi (Michałek, 2016).

W wyniku analizy nagrań wykonanych na potrzeby niniejszej pracy nie stwierdzono, aby zastosowanie funkcji pauzy w trakcie rejestracji powodowało zaburzenia struktury danych w plikach AMR.

## 5. Oprogramowanie do analizy plików AMR

Na podstawie opisanych wyżej wyników badań struktury plików w formacie AMR opracowano własne algorytmy i program komputerowy do automatycznej analizy tego typu plików dźwiękowych. Program ten skonstruowano i zaimplementowano w środowisku obliczeniowym MATLAB. Jego schemat blokowy przedstawiono na rysunku 6.

Ogólna zasada działania algorytmu i oprogramowania do automatycznej analizy plików w formacie AMR jest następująca:

- 1) import wszystkich danych zapisanych w pliku dźwiękowym,
- 2) detekcja nagłówek pliku i rozpoznanie typu AMR,
- 3) detekcja nagłówek kolejnych ramek z zakodowanym dźwiękiem i określenie jego parametrów z wykorzystaniem wbudowanej bazy danych,
- 4) na podstawie odczytanych informacji z nagłówek analiza struktury pliku ramka po ramce oraz porównanie ze sobą parametrów wszystkich ramek,
- 5) obliczenie liczby danych dźwiękowych i dokładnego czasu trwania nagrania na podstawie odczytanych parametrów,
- 6) weryfikacja integralności pliku oraz detekcja ewentualnych zaburzeń struktury lub ramek o innych parametrach,
- 7) szczegółowy raport z wykonanej analizy pliku.

Istotnym elementem opisywanego oprogramowania jest przedstawienie raportu zawierającego wyniki automatycznej analizy pliku. Raport ten zawiera szczegółowe informacje o parametrach pliku i zapisanego dźwięku, dokładnym czasie trwania nagrania, wynikach analizy spójności struktury i liczbie poprawnych oraz ewentualnych niepoprawnych ramek w pliku AMR.

Na rysunku 7 przedstawiono przykładowy raport wygenerowany na podstawie analizy spójnego pliku o nazwie *Nagr002.amr* zarejestrowanego telefonem komórkowym Nokia 3100c.

Na rysunku 8 zaprezentowano wynik analizy pliku AMR wykonanego telefonem Nokia E52-1, w którym występują zaburzenia struktury danych.

## Wnioski

Wykonanie badań nad właściwościami dźwięku zapisanego w formacie AMR pozwoliło pozyskać i usystematyzować wiedzę o tego typu nagraniach i plikach oraz ocenić możliwości analizy ich autentyczności. Zapis nagrania w pozornie nieskomplikowanym formacie AMR umożliwia wykonanie badań mających na celu ocenę autentyczności takiego nagrania. Wyniki przeprowadzonych prac wykazały, że dla nagrań utrwalonych w tym formacie możliwe jest zastosowanie powszechnie uznanych metod badania autentyczności, przede wszystkim takich, jak analiza sygnału przydźwięku sieciowego oraz struktury danych w pliku. Metodami wspomagającymi mogą być analiza charakterystyk nagrania i audytywna. Wykorzystanie systemu operacyjnego Windows i popularnych programów do analizy plików multimedialnych dostarcza jedynie podstawowych i często nieprecyzyjnych informacji o plikach zapisanych w formacie AMR. Na podstawie wyników z przeprowadzonych analiz struktur plików AMR zaprojektowano i wykonano własne oprogramowanie do automatycznej analizy tego typu plików pracujące w środowisku obliczeniowym MATLAB. Oprogramowanie to wyznacza poprawne parametry oraz wykonuje analizę integralności pliku i detekcję ewentualnych zaburzeń struktury. Zgromadzone i przeanalizowane nagrania testowe oraz wykonane oprogramowanie do automatycznej analizy plików AMR stanowią użyteczne narzędzia podczas badań autentyczności nagrań dowodowych, jakie wykonuje się w związku z działalnością opiniodawczą w Pracowni Analizy Mowy i Nagrań Instytutu Ekspertyz Sądowych.