Best Practice Manual for Digital Audio Authenticity Analysis

CONTENTS

ACKNOWLEDGEMENTS

Anna Bartle (Metropolitan Police Service, United Kingdom), Alexander G. Boyarov (The Russian Federal Center of Forensic Science of the Ministry of Justice, Russia), Dagmar Boss (Bavarian State Criminal Police Office, Forensic Science Institute, Germany), Luca Cuccovillo (Fraunhofer Institute for Digital Media Technology IDMT, Germany), Catalin Grigoras (National Center for Media Forensics, University of Colorado Denver, CO, USA), Marcin Michałek (Institute of Forensic Research, Poland), Dan Nyberg (Swedish National Forensic Centre, Sweden), and all other contributors are gratefully thanked for their invaluable contributions to the preparation of this guidance document.

1 AIMS

This BPM aims to provide a framework for procedures, quality principles, training processes and approaches to the forensic examination. This BPM can be used by Member laboratories of ENFSI, by other forensic science laboratories and by forensics experts to establish and maintain working practices in the field of forensic digital audio authenticity analysis that will deliver reliable results, maximize the quality of the information obtained, and produce robust evidence and unbiased conclusions. The use of consistent methodology and the production of more comparable results will facilitate interchange of data between laboratories.

The term BPM is used to reflect the scientifically accepted practices at the time of creation. The term BPM does not imply that the practices laid out in this manual are the only good practices used in the forensic field. In this series of ENFSI Practice Manuals the term BPM has been maintained for reasons of continuity and recognition.

2 SCOPE

This BPM addresses the forensic authenticity analysis of digital audio recordings. It provides recommendations concerning required resources, available scientifically validated methods and applicability thereof, quality assurance, handling of the recording under analysis, and interpretation guidelines. This document does not describe the methods for evidence gathering from digital media storage or equivalent. It assumes that the forensic principles for evidence handling are followed and addresses all necessary operations starting from when the audio recording is submitted together with an examination request. This BPM does not provide example analysis reports since they may vary considerably according to the laws and to the capabilities of the facility performing the analysis.

3 DEFINITIONS AND TERMS

Audio authenticity analysis – Act of providing an assessment about the evidence having characteristics compatible with an authentic digital recording or not.

Authentic digital recording – As applied to audio recordings, a continuous recording made simultaneously with the acoustic events, in a manner fully and completely consistent with the

method of recording, stored on a recoverable digital format, and which is free from unexplainable artefacts or discontinuities.

Audio file format – An organized structure for storing an audio recording in a file.

Contextual information – Additional information provided about the evidence, specifying the recording conditions and methods, e.g., date, time and place of the recording, type and configuration of the device and software involved, presence of known interruptions (pressing pause button, receiving an incoming call).

Cryptographic hashing functions – Publicly known algorithms used to map data of arbitrary size to a single fixed-length sequence of bits, referred to as hash or hash value. These values can be used, e.g., to substantiate the integrity of digital evidence or for comparisons against sets of known values. Hash computation must be efficient, deterministic, unforgeable, and grant low collision probability.

Digital audio recordings – Representation of audio signals by means of a set of numerical values, each value representing a discrete time instant.

Electric Network Frequency – Instantaneous frequency of the electric network, varying smoothly and randomly around the nominal operative value (50Hz in continental Europe).

Metadata – Data containing information about a file. As applied to audio recordings, it may store information about, e.g., audio parameters, codecs, dates and times, hardware or software involved.

Wiped storage media – Storage media which has been processed to erase any trace left by previous files which have been stored on it, e.g., by overwriting its content with random bits.

4 RESOURCES

4.1 Personnel

Personnel should have received specific forensic training in the field of audio forensics. Example appropriate trainings are:

- Laboratory in-house training
- Training from university or equivalent
- Training from external certified organization

Which training types are allowed and recognized may vary according to the specific legislation and may also vary between laboratories.

4.2 Equipment

Suitable equipment is required to perform proper audio analyses:

- Computer and high-quality audio card with audio resolution ranging from 8 kHz — 48 kHz, 16 bit, stereo
- High quality headphones with full frequency resolution 20 Hz — 22 kHz, high quality loudspeakers are optional but recommended
- Computer software with the possibility to at least read and decode audio data in the following codecs or formats: AAC, MP3, WMA, WAV/AIFF, OGG, AMR.
- Computer software with the possibility to visualize the Waveform, Spectrum and Spectrogram of the audio files.

The choice of equipment is of primary importance: inappropriate equipment may significantly degrade the quality of the forensic analysis. See for example [6] on the effects of peripheral stimuli and equipment used on Speech Intelligibility in Noise.

4.3 Reference Materials

Not applicable.

4.4 Accommodation and environmental conditions

Minimum requirements:

- Room with controlled noise level for in-house labs
- Noise cancelling headphones or equivalent for mobile labs

4.5 Materials and Reagents

Not applicable.

5 METHODS

5.1 Principles of Audio Authenticity Analysis

Authenticity analysis of digital audio recordings is based on *traces* left within the recording during the recording process, and by other subsequent editing operations.

The first goal of the analysis is to *detect and identify* which of these traces can be retrieved from the audio recording, and to document their properties.

In a second step, the properties of the *retrievable traces* are analysed, to determine if they support or oppose the hypothesis that the recording has been modified.

5.1.1 Recording traces

The majority of traces is left in the recording during the recording process, as depicted in Figure 1.

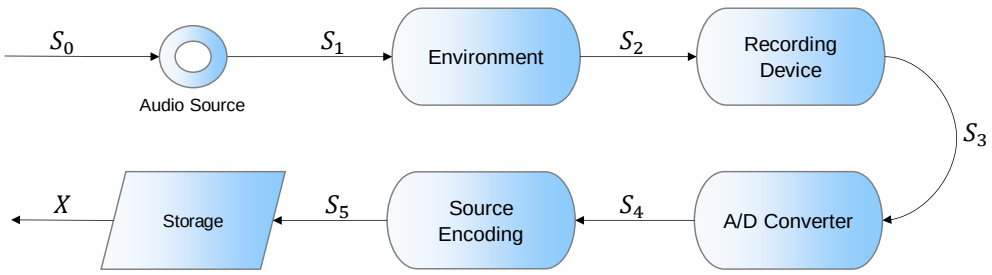

Figure 1: Recording process flow

In a first step, the speaker's thoughts S_0 are converted into the original speech signal.

The speech S_1 is then propagated through the environment. During this acoustic propagation, S_1 can be modified by reverberations and mixed with random, periodic or harmonic environmental noise.

This complex mixture S_2 is then converted to an electrical signal by the microphone (transducer). During the recording, it is modified by the microphone frequency response, influenced by microphone thermal noise, and may obtain additional spurious traces such as a DC-offset or an electrical network frequency (ENF) component.

The analogue electrical signal S_3 is then converted into its digital representation, according to the specific bit-depth and sampling frequency used by the analogue-to-digital converter.

The digital representation S_4 is then encoded using either a lossy or lossless scheme, thus acquiring encoding artefacts, bitrate, mono or stereo mode, and any other parameter specific to the scheme.

Lastly, the encoded signal S_5 is stored together with any related metadata, on the device storage as the original evidence X.

5.1.2 Traces of post-processing

A second class of traces are present whenever the recording has gone through editing or other operations due to human intervention, e.g.:

- Inter-sample dependencies left by digital resampling of the original content, in which the output contains correlations between neighbouring audio samples not compatible with the random nature of the input signal.
- Double-encoding artefacts, such as musical noise, due to several lossy-encoding schemes applied one after another
- Replicated time intervals, i.e., time intervals in which the content is perfectly identical which does not happen in natural speech recordings. The presence of such replicas is a sign of human post-processing of the initial recording.

The main and most important characteristic of such traces is that they *cannot be attributed to any part of the purported (claimed) recording process*. Hereon, we will refer to this class of traces with the term “traces of post-processing”.

Operations which modify the content of the recording, such as deleting a portion of the file, copy-pasting a time interval coming from the same file, splicing content from a different recording, may also generate *discontinuities* in the recording traces.

Other operations, as the aforementioned resampling and double encoding, may not affect the content of the audio recording, but may mask other editing or manipulation signs. Thus, the importance given to the *detection* of traces of post-processing.

5.1.3 Hypothesis Testing

Audio authenticity analysis, ideally, should be able to answer the following question:

Is the evidence under analysis an authentic recording[1] or not?

In practical casework, authenticity analysis rather consists on *supporting/rejecting* the hypothesis that evidence is an authentic recording, based on the characteristics of the traces within the recording and the available contextual information.

If we consider the two following alternative propositions:

- H0: the evidence presents traces supporting the hypothesis that the recording is authentic
- H1: the evidence presents traces supporting the hypothesis that the recording is not authentic

it should be clear that the goal of authenticity analysis is *not* to state which proposition is the correct one, but to *quantify* which hypothesis is the more likely, and how strong (or weak) this support is. It is possible to have *no support* for either case.

Proposition H1 can be expressed also by means of the following statement, which we can consider the fundamental principle of audio authenticity analysis:

If any recoding trace is inconsistent at any point in the file
OR
If unexplainable post-processing traces are present
} The recording is not authentic

The above principle implies that the analysis should not strive for explicit authentication but rather focus on post-processing detection. That implies the following:

1. *Every* detected recording trace – e.g., microphone frequency response, encoding (scheme, bitrate, mono or stereo-mode), ENF, reverberation, DC-offset, bit-depth, metadata –- should be checked for inconsistencies.
2. *Every* detected recording trace should be checked for inconsistencies with any existing reference audio recording, as well as with any contextual information regarding the acquisition process.
3. *Any* detected post-processing trace should be thoroughly documented and compared with the provided contextual information.

In the following sections we are going to introduce several forensics methods, all of which are related to the fundamental principle above, and to describe how these methods should be applied. See section 12 for evaluation and interpretation.

[1] The definition of authentic recording is provided in section **Fel! Hittar inte referenskälla.**

5.2 Methods classification

Methods for audio forensics analysis can be divided in two main classes:

1. *Informed analysis methods*, focused on the detection of inconsistencies between traces left by the recording process and contextual information regarding the recording process of the evidence, possibly with the help of reference recordings.
2. *Blind analysis methods*, relying solely on the content, focused on the detection of inconsistencies of traces left by the recording process within the file, and on the detection of traces that may have been left by editing.

Informed analysis methods can be described in terms of:

a) Which recording trace is addressed
b) Which contextual information has been used for building the ground truth

Blind analysis methods should be categorized according to the traces and their *properties*:

- Continuity of time-variant traces
  Methods verifying the continuity of: speech signal, environmental noise signals, electric network frequency
- Invariability of time-invariant traces:
  Methods verifying the lack of changes of: reverberations, constant background noise, DC-offset, microphone frequency response, microphone thermal noise, bit-depth, cut-off frequency, encoding artefacts and any noises/traces caused by the electronic elements
- Invariability of periodic traces:
  - Methods verifying the periodicity of: periodic environmental noise, periodic encoding artefacts.
- Detection of traces of post-processing:
  - Methods detecting: inter-sample dependencies, double encoding traces, replicated time intervals, abnormal distribution of quantization levels.

With this categorization, every method can thus be described in terms of:

i. Which trace is addressed
ii. Which property of the trace is checked

See Figure 2 and Figure 3 for examples of methods which can be used in each category of traces.

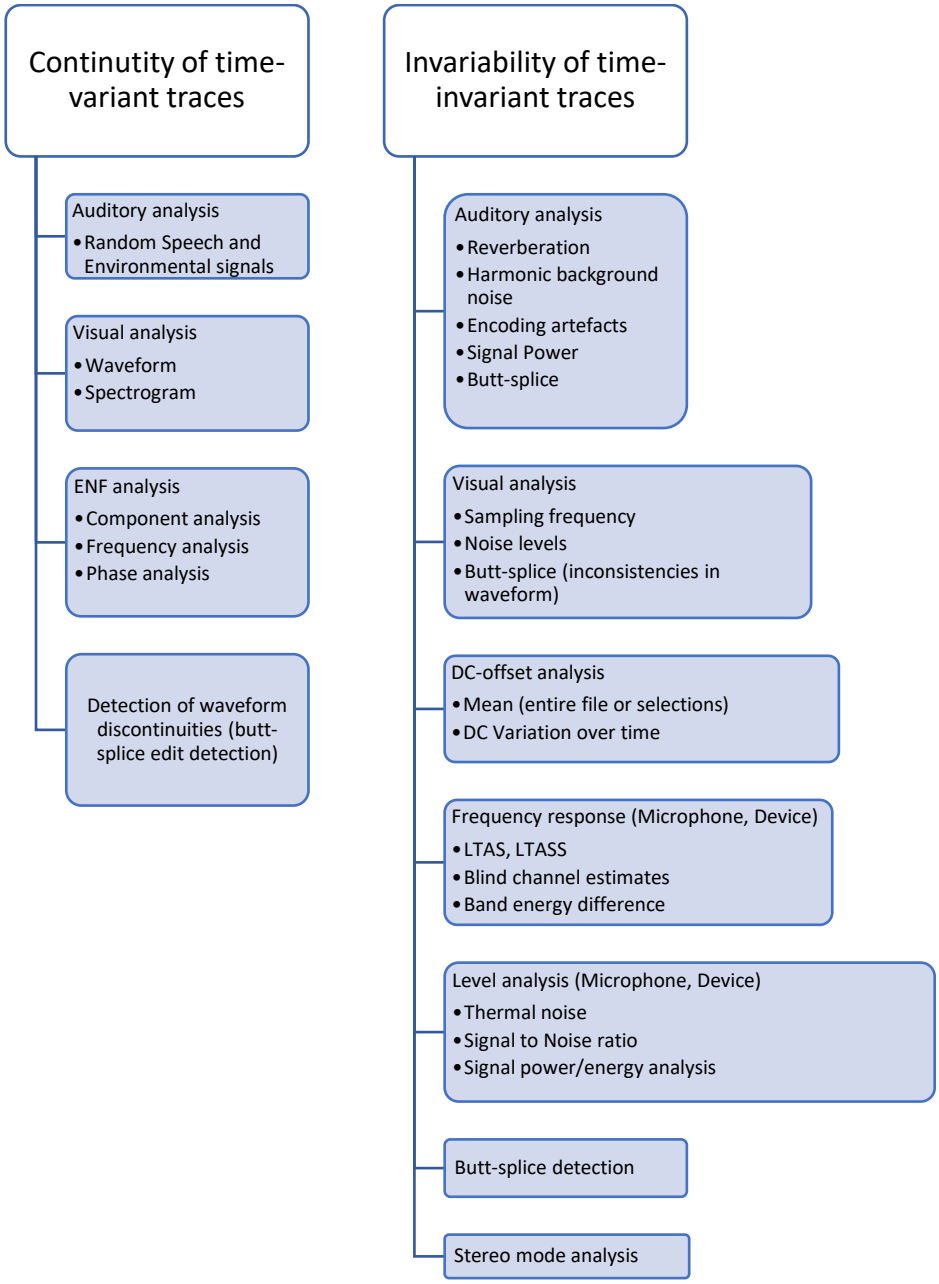

Figure 2: Available methods categorized according to the trace properties.

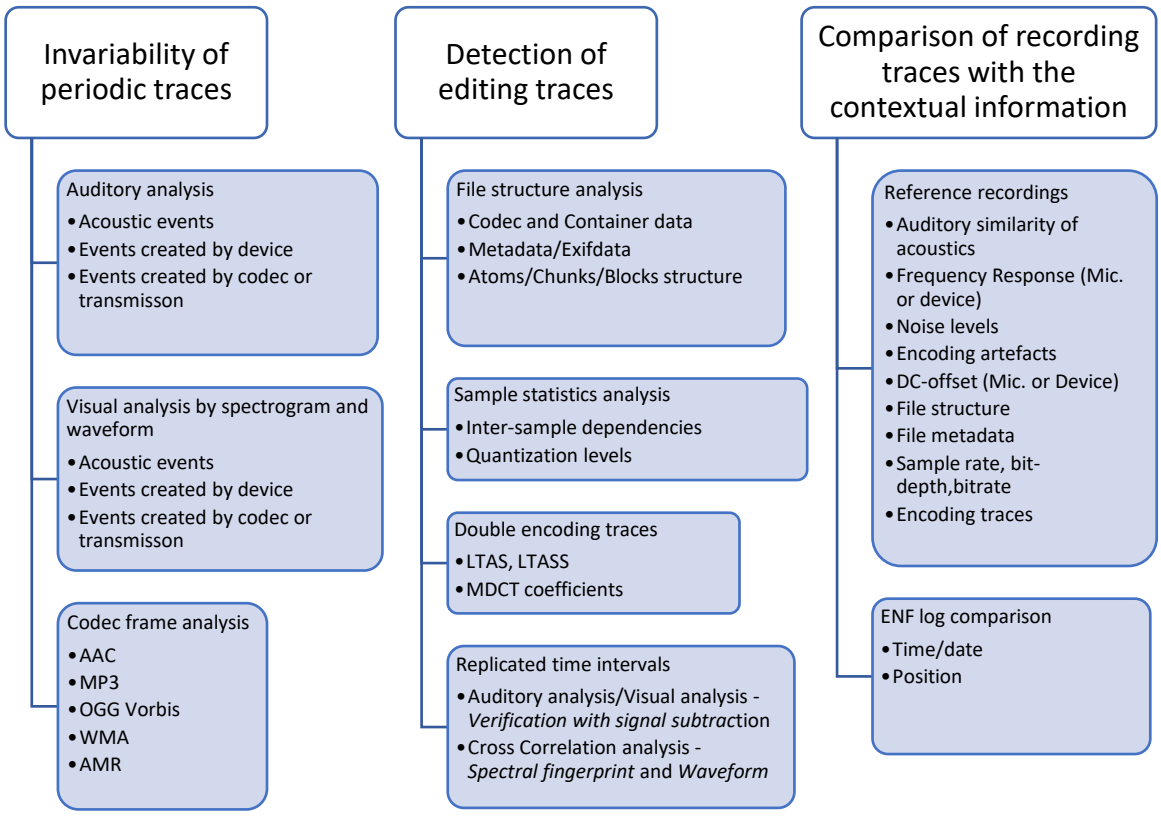

Figure 3: Available methods categorized according to the trace properties, continued

5.3 Laboratory and Case specific framework

The choice in terms of the specific implementation of each proposed method is left to each laboratory, according to the available tools, technologies, and knowledge. It is recommended that audio laboratories maintain a chain-of-custody for evidential material whenever possible.

We encourage forensic practitioners to follow peer-reviewed methods, widely accepted by the scientific community, and thoroughly evaluated *within the laboratory*.

Some of the technology in the research field may not yet be ready for application in real-life conditions: The choice on which *scientifically validated* methods should be adopted is left to each lab, which is responsible of its choices.

The implementation of chosen methods is also dependent on the specifics of the examination

- "traces should be consistent *within* the recording," with the goal to find out whether there are any unexplained inconsistencies within the audio recording

And / or

- "traces should be consistent *with the contextual information* provided concerning the recording," with the goal to find out whether there are any inconsistencies with the statements provided concerning the evidence

The latter requires detailed information regarding the contextual aspects and purported provenance of the recording for example: recording device, recording software, any additional hardware (e.g. headset/headphones); time, location, speakers' identities; known events which may create *expected* artifacts – e.g., phone calls, TV switching channels etc. See Section 9.

Both can be valid approaches if allowed by the laboratory and juridical system.

5.4 Method descriptions

5.4.1 Continuity of time-variant traces

When looking at the continuity of time-variant traces, the focus of the auditory analysis is to listen for inconsistencies in the signal. For example, unnatural fade-in/fade-outs and abrupt changes in a word or in an environmental sound such as a passing car or café ambiance. Analysis of waveform and spectrogram can be used to verify and supplement the findings from the auditory analysis. It is beneficial to compare the findings from the auditory and visual analyses with provided contextual information. Apparent inconsistencies can in many cases be explained by this additional contextual information. For example, changes in environmental sounds could be due to opening or closing a window, a door or by changing rooms during a recording. For more details see [7][8] and regarding the limitations of auditory and visual analyses see [9].

In addition to the intended signal the recording equipment may also capture the Electric Network Frequency (ENF) signal; that is, an alternating current power hum which fluctuates smoothly around its nominal value of 50 or 60 Hz (according to the location in which the recording is made). This may happen due to induction in the microphone itself, the cables involved, or the internal electronics of the recording device. When this happens, the recording contains a series of harmonic tones the fundamental of which is the ENF [45][46][47].

Due to the ENF variations being very slow and well defined in their magnitudes, deletions or addition of content may create discontinuities in the frequency [48][49] and / or the phase trajectories [50][51] of the ENF component extracted from the recording in question. As these discontinuities sometimes also happen on unaltered recordings, due to local impulse noises on the network, discontinuity analysis should be performed together with at least one other analysis – for example, determining whether the ENF matches the purported recording time [52][53][54].

When the recording process described in Figure 1 occurs in a single electrical network, then the original digital recording is expected to contain one series of ENF harmonic tones or none. More than one ENF series can indicate that analogue copying of the original signal may have taken place, with the additional ENF trace having been captured during the copying process.

5.4.2 Invariability of time-invariant traces

Here the auditory analysis involves listening for:

- Unnatural changes in reverberation
- Unnatural changes in background sounds
- Unnatural changes in encoding artefacts
- Unexpected differences in sound levels including clicks, pops and level changes of sound sources.

The visual analysis involves looking for unexpected changes in spectral characteristics and waveform such as:

- Changes in cut-off frequency (which may have occurred due to recordings with different sample rates being spliced together)
- Changes in noise levels
- Sudden/abrupt changes in the waveform, e.g., butt-splices
- Sudden and unexplained changes in signal level

In original, unaltered recordings, the DC-offset of a device has a well-defined mean and finite standard deviation which can be estimated using the whole recording duration [20][21][22]. Intervals in which the local DC-offset is very different to the global one, may have been caused by portions of file having been added by editing. Such occurrences should also be examined and verified by auditory analysis, since DC-offset mismatches may generate audible clicks.

The frequency response of a recording device does not change abruptly. Hence, methods which examine features related to the microphone or device frequency response [38][39][40] may be used to identify possible additions of content recorded on a different device, or to make comparisons with test recordings made on a declared device.

Power and level analysis as described in [9] can be used to look for both abrupt changes and more gradual changes in sections in the signals power and level. Such changes may be introduced, among others, by deletion of material, by changes in the noise-floor happening when different recordings are edited together, but also by drop-outs or sudden changes of the recording conditions.

The encoding parameters of a recording do not change over the file length. For uncompressed WAV files, this implies that the real bit-depth, the PCM coding and mono or stereo mode of the base signal should remain constant throughout the file duration [24][26]. For compressed (e.g. MP3, AAC) files, the same would be true for the bitrate (in case of CBR), encoding quality (in case of VBR), cut-off frequency of the lowpass filter and mono or stereo mode, which can be recovered from the compressed domain, as well as after decoding the file [27][28][29][30][31][32].

Butt-spliced edits or other discontinuities occurring between one sample and the next may be detected in PCM encoded files by examining the 1st or 2nd order differentials of sample values against time [23]. Changes in sample value that are dissimilar to the rate of change of sample values in the immediately surrounding audio may result in an impulse in the plot of the difference signal. If perceptual encoding is applied after editing, such discontinuities are no longer detectable with the method in [23]. It is important to acknowledge that discontinuities are not necessarily caused by editing, so evidence of discontinuities is not evidence of editing.

5.4.3 Invariability of periodic traces

The focus of the auditory and visual analyses in this step is to listen and look for sudden changes in traces coming from the transmission, device or background sounds, such as unexpected changes in periodic dropouts created in the signal from transmission or encoding process, or unexpected changes of the rate of a wall clock in the room (or similar) that can be heard in the background of the recording.

Periodic traces from lossy encoding are also embedded in a file. In the case of transform-based codecs such as MP3, AAC, WMA and Ogg Vorbis, this periodicity is reflected by the framing grid offset, which can be estimated and analysed using the inverse decoding paradigm [33][34][35]. In the case of speech codecs using CELP or LPC as AMR-NB, FR, HR, EFR, the

residuals obtained by re-encoding the file with the same scheme exhibit a strong periodicity equal to the length of the base block used for LP analysis [36].

5.4.4 Detection of traces of post-processing

Analysis of the file structure and metadata is an important method used in authenticity investigations of digital audio recordings. This method is based on the visualization of the file structure in the hexadecimal and ASCII representation using hexadecimal viewing software, and on searching for significant information about a file condition. For example information from post-processing software may sometimes be seen in the metadata [10][11][12][13][14][15][16][17][18][19].

The presence of time intervals in which the content is perfectly identical to the one present in other sections is a strong sign of human post-processing of the initial recording. The occurrence of such intervals, which may correspond, e.g., to single words or short utterances, can be identified via auditory and visual analyses. The presence of a replicated portion may be verified using signal subtraction: if the signal is cancelled out by this process a copy/paste might have been made for that particular part of the signal.

Auditory analysis can also be used as a confirmation tool in order to verify a finding from an automatic method, e.g., copy-move forgery detection methods based on correlation analysis [62] or audio fingerprinting [63].

Digital resampling is a process for which the sampling frequency of the file is converted from the original value fs_{old} to a new value fs_{new}. The relation between the two frequencies can often be expressed by means of the formula $fs_{new} = (P/Q) \cdot fs_{old}$, with P and Q being coprime. E.g., up-sampling from 16kHz to 24kHz can be written as $24kHz = (3/2) \cdot 16kHz$, and down-sampling from 44.1 kHz to 8kHz can be written as $8kHz = (80/441) \cdot 44.1kHz$.

Whenever digital resampling by a rational factor P/Q is applied to a recording, with P and Q being coprime, re-sampling creates periodic inter-sample dependencies appearing within blocks of Q samples. This periodicity can be identified for uncompressed files both in the case of up-sampling with $P/Q > 1$ [67] and in the case of down-sampling with $P/Q < 1$ [68] even if in this second case the detection accuracy is lower due to the task being more challenging.

Double encoding effects appear whenever a lossy-encoded file is firstly decoded, and then re-encoded using the same or a different format. At the time of writing, only the case of double encoding using the same scheme has been investigated, and successfully applied to both MP3 and AAC. The detection was performed by comparing an artificially simulated single-encoded version of the input file with the actual evidence, to determine the presence of double encoding artefacts [69][70][71].

A change in amplitude (i.e., digitally applied gain or attenuation) of the audio content may lead to anomalies in the quantization levels used in the recording. The number of quantization levels would be that dictated by the encoder, but their distribution may be affected in a processed recording, resulting in visible gaps and periodicities within the histogram of the quantisation levels [24].

5.4.5 Comparison of recording traces with the contextual information

If the location (i.e., the environment in which the recording took place) is known, then a comparison of the acoustical properties may aid the analysis/investigation. At the time of writing, no automatic environment classification analysis has been proposed able to cope with real case scenarios, with or without reference recordings.

Methods such as estimating the decay rate (e.g. RT60 parameter) of late reverberations [64] or applying de-reverberation techniques to then build a profile of the reverberant signal [65] are considered to be unreliable at the time of writing [66].

If a reference recording can be produced in the alleged environment using the alleged recording device and setup (including locations of microphone and acoustic sources), a comparison via critical listening can be carried out.

If the purported / alleged location in which the recording was made is known, and the evidence contains an ENF-signal of adequate quality, the nominal frequency (50/60 Hz) of the ENF should match the one used by the electric network in the provided location. A fine-grained localization, however, is not evidentially reliable at the time of writing [55][56][57].

If the evidence contains an ENF-signal of adequate quality and length, *ENF temporal pattern matching* can be performed to determine the time of recording.

ENF temporal pattern matching requires a ENF database which has been properly created and maintained [46][51][58][59]. If an adequate database is available, the frequency trajectory of the ENF component extracted from the evidence can be compared to it according to [45][46][47][51][60][61]. Some ENF databases can also be accessed on demand from trusted sources.

To minimize errors, time and frequency resolutions used for extracting the ENF should match the ones used in the database. Harmonics of the fundamental ENF may also be used to estimate the ENF signal trajectory on the evidence recording.

If the recording device alleged / purported to have made the recording is known, microphone analysis techniques can be used.

Microphone frequency response analysis may be used in portions of recordings in which the speech signal is predominant, to determine which, among a set of possible microphones / devices, is most likely to have be used to make the recording [38][39][40][41], and / or to provide information regarding the characteristics of the microphone / device used. Similarly, but focusing on nearly-silent unvoiced portions of the audio recording, *microphone thermal noise analysis* may be used [42][43][44].

A comparison of “general” frequency content can also be used in order to look at similarities to and differences from reference recordings. This comparison can be performed, e.g., by means of the LTAS and the LTASS features obtained for the evidence and the reference recordings [24][28]. It should be noted, however, that these features are influenced not only by the frequency response of the microphone / recording device, but also by other factors such as the encoding parameters or the signal content. Hence the recording conditions must be known and documented as far as possible, to avoid errors.

Comparison of *microphone DC-offset analysis* against reference recordings can also be used [20][21][22].

To avoid any analysis bias, the direct comparison of two recordings in terms of microphone traces should be performed only when the recording conditions are similar, e.g., with similar environment, presence or absence of speech, device approximate signal amplitude and encoding settings. The features used for the comparison must also be obtained using the same analysis settings (e.g., window length, shape, hop-size, type and number of filter-banks) in both the evidence and the reference file. All settings should be thoroughly documented, for reproducibility.

Frequency response and thermal noise analysis may support the hypothesis of a specific recording device being involved or may be used to exclude specific devices from a list of candidates. DC-offset analysis, however, should never be used to support the usage of a specific device, but only to exclude specific devices from a list of candidates [20][21][22].

If the purported model of the audio recorder and the recording software and version are known, encoding analysis can be performed.

If the submitted evidence is provided as WAV file, *inverse decoding* can be used to check that there are no unexpected traces of lossy encoding [26][30][32][36]. If the evidence is provided as compressed (e.g., MP3, M4A) file, *statistical encoding analysis* can be used to verify the 'real' bitrate derived from the baseline signal [27][28][37], in addition to the one declared on the container.

In some cases, recorders may perform lossy encoding to then store the file in an uncompressed format – thus creating "expected" traces of lossy encoding in WAV files stored on the device. Other recorders may store an encoded file using a higher or lower bitrate than the "real" one used during the recording – thus creating "expected" traces of double-encoding in the compressed files stored on the device. It is therefore important, whenever possible, to make test recordings if the purported software / device are known, and to consider the whole range of options and settings available on the purported recording software / device.

If the purported model of the audio recorder, the recording software and version, as well as the recording settings are known, metadata and file structure analysis can be performed.

Given the input evidence, *metadata and file structure analysis* can be performed to identify which metadata are present on the file, and in which order they are present. Metadata can be categorized into functional (necessary for a correct file playback, mostly dependent on the format used for storage), library related (introduced by the specific encoding library) and software related (introduced by the specific software used for creating the file). These three categories should be considered in conjunction with one another, and it is good practice to store expected values for these metadata and file structure in reference databases [12].

Example metadata and file structure analyses can be found for WAV files [10][13], MP3 files [10][14], WMA files [10][15], M4A files [11][17], and AMR files [16]. Although these publications may be used as reference, it is important to underline that the consistency in metadata between questioned and reference recordings is insufficient to claim the absence of any modification, since metadata can also be edited as part of the tampering process.

If the purported model of the audio recorder, the recording software and version, as well as the purported recording settings are known, and the submitted file is in uncompressed WAV format, *quantization level analysis* can be performed.

During the A/D conversion stage, the input analogue signal is digitised using a specific hardware related bit-depth. Bit-depths offered by A/D converters do not always match the bit depth of WAV files: the ones stored using 16-bit linear PCM may thus contain signals quantized with 11,12,14 bits. This *real* lower bit-depth can be retrieved and compared to the one found for reference files [24].

It is important to underline that the consistency in quantization levels between questioned and reference recordings is insufficient to claim the absence of any modification, since they can be edited by expert audio engineers as part of the tampering process.

5.5 Remarks

5.5.1 Dealing with discontinuous recordings

According to the national legislations, some cases may require the analysis of recordings which are *known* to be discontinuous from the contextual information: such cases include, e.g., a phone call interrupting the recording by a mobile phone, or the user pressing a pause button on a hand-held recorder.

In these cases, the recording process was not continuous, but the continuity of fragments between *documented or explainable* discontinuities may still be examined. Furthermore, depending on the recording software and equipment, discontinuities (e.g., pauses) may be indicated by specific signal traces and/or metadata in the file under examination. If this is the case, the analysis would proceed as normal, but in addition the declared discontinuities would be tested to ensure that they are consistent with the declared sequence of events.

5.5.2 Global / local analysis methods

Most methods described in Section 5.4 address signals with a finite specific length, i.e., they are described in terms of an analysis window which may or may not span the whole file length. Whenever this is the case, then the methods can be used for both global and local analysis.

Global analysis, performed on the whole file length, may be used to check the consistency of the detected traces with the contextual information provided on the content. E.g., it may be used to perform a first screening based on the ENF traces being compatible with the recording time/date, or on the consistency of quantization levels with the container bit-depth.

Local analysis, performed on consecutive analysis windows, may be used instead to check the consistency of the detected traces within the recording and / or to detect artefacts supporting a hypothesis of content modification. E.g., the analysis may look for ENF phase discontinuities on silent portions of the file, or whether the distribution of used quantization levels varies through the file duration.

Where applicable minimum requirements for the length of the analysis windows (and sometimes, of the best operative length of such windows) are stated in the previously cited publications. In case of doubts, practitioners are encouraged to contact the authors for clarification, rather than risking misuse of the methods.

6 VALIDATION AND ESTIMATION OF UNCERTAINTY OF MEASUREMENTS

Validation can be done for each one of the implemented measurement methods. As a minimum three conditions should be used for each validation study. The dataset for validation

should always mirror the casework relevant for the laboratory. For example, for the validation of a method for detecting recompression of an audio file, the conditions maybe: file with no recompression as baseline, recompression to the same resolution, recompression to a lower resolution and a higher resolution.

For methods based on human perception, i.e., auditory and visual analysis (waveform/spectrogram), it is strongly recommended that at least two examiners conduct independent analyses. The results should be compared and summarized.

7 QUALITY ASSURANCE

It is recommended that laboratories participate in any suitable available proficiency testing or collaborative exercise every three years. If this is not possible it is recommended that interlaboratory exercises are conducted with two or more laboratories every other year.

8 HANDLING ITEMS

Practitioners should always create and work on a working copy of the audio evidence, and not on the original submitted copy. To ensure that an error-free 1:1 copy of the evidence has been obtained, cryptographic hash functions, for example SHA-2 or SHA-3 should be used. Some software needs an uncompressed file format to conduct an analysis. If so, the conversion algorithm used should be validated to ensure the integrity of the audio content is maintained.

If the evidence digital recording system is provided, then a forensic image of the memory content should be created and the analysis run on a copy of the evidence files as stated above. If the evidence digital recording system has removable media, then it should be replaced with a similar one for the creation of test or reference recordings. Access to the evidence memory protected by write blockers where it is technically possible to do so.

9 INITIAL ASSESSMENT

During the initial assessment of material and of the inquiry, contextual information as relevant to the case should be sought after, and ideally, should come from the person who purportedly made the recording. For example:

- In what context is the recording made? E.g. outside, covert recording, telemarketing etc.
- What equipment (both hardware and software) has been used during the recording? Both the information regarding the equipment and/or the actual physical set up of the equipment is very useful.
- Do the recording software and device implement algorithms for automatic gain adjustment, low-cut filter or automatic pausing during silence? Were these options active during the recording session?
- Was the recording session interrupted by phone calls?
- Did the recording session take place with a static setup, or was the microphone hand-held and moved across the environment?
- At what time/date has the recording been made?
- Is it a first generation recording or was it transferred or re-encoded?
- Was the signal processed, e.g., for speech enhancement?

The main purpose for asking these questions is to attain a foundation of information and data which may be supported or refuted through the examinations undertaken. See Section 5.1.3.

Furthermore, when relevant, any specific allegations regarding tampering should be specifically specified by the submitting party.

During the initial assessment the quality and quantity of the questioned recordings should always be assessed, in order to determine which methods can be used in the investigation. For example, if the signal to noise ratio is very low then a particular method for copy/clone detection might not be possible to use.

10 PRIORITIZATION AND SEQUENCE OF EXAMINATIONS

Not applicable.

11 RECONSTRUCTION

When possible, reference recordings should be created using reference equipment (same brand, model, firmware and application, file format and settings as stated in the inquiry and data derived from the case). If no reference equipment is available and reference recordings need to be produced on the evidence device, a forensic image of the entire evidence equipment's memory should be taken (where technically possible) before recording the references. *If possible, the production of reference recordings on the evidence equipment should be avoided, unless the equipment records to removeable media.* If the evidence equipment uses removable media, the media containing the questioned recordings should be removed and new or wiped media used for the reference recordings.

Practitioners should try to create reference recordings containing audio material similar to that on the questioned recording. For example, if the recording contains equipment handling noise or wind noise, sounds from a café or traffic, the reference recording should be made with similar sounds where possible. This aids comparison of the recordings.

12 EVALUATION AND INTERPRETATION

It is recommended that the hypotheses addressed should be formulated based on the inquiry by the client and the contextual information given. As stated in Section 5.1 and 5.1.3 the main goal is to retrieve, document and analyse all available traces.

All the results collected from the investigation should be taken into consideration and the level of support for each hypothesis should be stated. The latter can be done and presented in many different ways, in tables or plain text etc. The conclusions may be expressed according to a numerical scale and / or a verbal scale.

When formulating the conclusion, it is of utmost importance that the conditions and limitations on which the conclusion is based are stated.

13 PRESENTATION OF EVIDENCE

The overriding duty of those providing expert testimony is to the court and to the administration of justice. As such, the outcome of an evidence analysis should be provided with honesty, integrity, objectivity and impartiality.

Evidence can be presented to the court either orally or in writing. Presentation of evidence should clearly state the results of any evaluation and interpretation of the examination. Written reports should include all the relevant information in a clear, concise, structured and unambiguous manner as required by the relevant legal process. Written reports must be peer reviewed.

Expert- witnesses should resist responding to questions that take them outside their field of expertise unless specifically directed by the court, and even then, a declaration as to the limitations of their expertise should be made.

It is recommended that the questioned audio material should be played back when presenting the evidence to the court. The equipment used for playback should be adapted to the special acoustic conditions of a courtroom (good loudspeakers or headphones).

14 HEALTH AND SAFETY

Not applicable.

15 REFERENCES

General Information

[1] ENFSI. (2005). “Code of conduct BRD-GEN-003 (002).”
[2] SWGDE. (2016). “Digital & multimedia evidence glossary (3.0).”
[3] ASTM. (2013). “E2916 standard terminology for digital and multimedia evidence examination.”
[4] ENFSI. (2015). “Best practice manual for the forensic examination of digital technology ENFSI-BPM-FIT-01 (1.0).”
[5] SWGDE. (2017). “Best practices for digital audio authentication (1.2).”

Forensic Audio Authentication

[6] D. Bergfeld and K. Junte, “The effects of peripheral stimuli and equipment used on speech intelligibility in noise,” in *AES International Conference on Audio Forensics*, Arlington, VA, USA, 2017.
[7] B. E. Koenig and D. S. Lacey, “Forensic authentication of digital audio recordings,” *Journal of the Audio Engineering Society*, vol. 57, no. 9, pp. 662–695, 2009.
[8] E. B. Brixen, “Techniques for the authentication of digital audio recordings,” in *112th AES Convention*, Vienna, Austria, 2007.
[9] C. Grigoras, D. Rappaport, and J. M. Smith, “Analytical framework for digital audio authentication,” in *AES International Conference on Audio Forensics*, Denver, CO, USA, 2012.

Structure and Format Analysis

[10] C. Grigoras and J. M. Smith, “Large scale test of digital audio file structure and format for forensic analysis,” in *AES International Conference on Audio Forensics*, Arlington, VA, USA, 2017.
[11] J. M. Smith, D. S. Lacey, B. E. Koenig, and C. Grigoras, “Triage approach for the forensic analysis of apple iOS audio files recorded using the “Voice Memos” app,” in *AES International Conference on Audio Forensics*, Arlington, VA, USA, 2017.
[12] M. Michałek, “Test audio recordings and their use in authenticity examinations. Database of properties of digital audio recorders and recordings,” *Problems of Forensic Sciences*, vol. 105, pp. 355–369, 2016.
[13] B. E. Koenig and D. S. Lacey, “Forensic authenticity analyses of the metadata in re-encoded WAV files,” in *AES International Conference on Audio Forensics*, London, United Kingdom, 2014.
[14] B. E. Koenig, D. S. Lacey, and C. E. Reimond, “Selected characteristics of MP3 files re-encoded with audio editing software,” *Journal of Forensic Identification*, vol. 64, no. 3, pp. 304–321, 2014.
[15] B. E. Koenig and D. S. Lacey, “Forensic authenticity analyses of the header data in re-encoded WMA files from small Olympus audio recorders,” *Journal of the Audio Engineering Society*, vol. 60, no. 4, pp. 255–265, 2012.

[16] M. Michałek, “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*, vol. 109, pp. 27–42, 2017.
[17] M. Michałek, “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,” *Problems of Forensic Sciences*, vol. 115, pp. 241–261, 2018.
[18] M. Michałek, “The characteristics of popular audio recording applications installed on smartphones with an Android operating system in relation to forensic audio analyses,” *Problems of Forensic Sciences*, vol. 120, pp. 335–361, 2019.
[19] B. E. Koenig and D. S. Lacey, “Forensic authenticity analyses of the metadata in re-encoded iPhone M4A files,” in *AES International Conference on Audio Forensics*, Arlington, VA, USA, 2017.

Time Domain Analysis

[20] B. E. Koenig and D. S. Lacey, “The average direct current offset values for small digital audio recorders in an acoustically consistent environment,” *Journal of Forensic Sciences*, vol. 59, no. 4, pp. 960–966, 2014.
[21] B. E. Koenig, D. S. Lacey, C. Grigoras, S. G. Price, and J. M. Smith, “Evaluation of the average DC offset values for nine small digital audio recorders,” *Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 439–448, 2013.
[22] B. E. Koenig, D. S. Lacey, C. Grigoras, S. G. Price, and J. M. Smith, “Evaluation of the average DC offset values for nine small digital audio recorders,” in *AES International Conference on Audio Forensics*, Denver, CO, USA, 2012.
[23] A. J. Cooper, “Detecting butt-spliced edits in forensic digital audio recordings,” in *AES International Conference on Audio Forensics*, Hillerød, Denmark, 2010.

Encoding Traces Analysis

[24] C. Grigoras, “Statistical tools for multimedia forensics: Compression effects analysis,” in *AES International Conference on Audio Forensics*, Hillerød, Denmark, 2010.
[25] C. Grigoras and J. M. Smith, “Quantization level analysis for forensic media authentication,” in *AES International Conference on Audio Forensics*, London, United Kingdom, 2014.
[26] L. Cuccovillo and P. Aichroth, “Inverse decoding of PCM A-law and µ-law,” in AES *International Conference on Audio Forensics*, Porto, Portugal, 2019.
[27] D. Seichter, L. Cuccovillo, and P. Aichroth, “AAC encoding detection and bitrate estimation using a convolutional neural network,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 2069–2073.
[28] C. Grigoras and J. M. Smith, “Forensic analysis of AAC encoding on Apple iPhone Voice Memos recordings,” in *AES International Conference on Audio Forensics*, Porto, Portugal, 2019.
[29] A.G. Boyarov and I.S. Siparov, “Forensic Investigation of MP3 Audio Recordings,” *Theory and Practice of Forensic Science*, vol. 14, no. 4, pp. 125–136, 2019.
[30] R. Korycki, “Authenticity examination of lossy compressed digital audio recordings,” in *EAA Conference - Forum Acusticum*, Kraków, Poland, 2014.
[31] S. Moehrs, J. Herre, and R. Geiger, “Analysing decompressed audio with the “Inverse Decoder” – towards an operative algorithm,” in *112th AES Convention*, Munich, Germany, 2002.
[32] J. Herre and M. Schug, “Analysis of Decompressed Audio – the inverse decoder,” in *109th AES Convention*, Los Angeles, CA, USA, 2000.

[33] D. Gärtner, C. Dittmar, P. Aichroth, L. Cuccovillo, S. Mann, and G. Schuller, “Efficient cross-codec framing grid analysis for audio tampering detection,” in *136th AES Convention*, Berlin, Germany, 2014.
[34] R. Korycki, “Detection of montage in lossy compressed digital audio recordings,” *Archives of Acoustics*, vol. 39, no. 1, pp. 65–72, 2014.
[35] R. Yang, Z. Qu, and J. Huang, “Detecting digital audio forgeries by checking frame offsets,” in *ACM Workshop on Multimedia and Security*, Oxford, United Kingdom, 2008, pp. 21–26.
[36] J. Zhou, D. Garcia-Romero, and C. Y. Espy-Wilson, “Automatic speech codec identification with applications to tampering detection of speech recordings,” in *ISCA Annual Conference (INTERSPEECH)*, Florence, Italy, 2011, pp. 2533–2536.
[37] R. Korycki, “Authenticity investigation of digital audio recorded as MP3 files,” *Issues of Forensic Science*, vol. 283, no. 1, pp. 54–67, 2014.

Microphone Analysis

(frequency response)

[38] L. Cuccovillo and P. Aichroth, “Open-set microphone classification via blind channel analysis,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 2069–2073.
[39] L. Cuccovillo, S. Mann, M. Tagliasacchi, and P. Aichroth, “Audio tampering detection via microphone classification,” in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Pula, Italy, 2013, pp. 177–182.
[40] D. Luo, P. Korus, and J. Huang, “Band energy difference for source attribution in audio forensics,” *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 9, pp. 2179–2189, 2018.
[41] C. Krätzer, A. Oermann, J. Dittmann, and A. Lang, “Digital audio forensics: A first practical evaluation on microphone and environment classification,” in *ACM Workshop on Multimedia and Security*, Dallas, TX, USA, 2007, pp. 63–74.

(thermal noise)

[42] R. Buchholz, C. Krätzer, and J. Dittmann, “Microphone classification using fourier coefficients,” in *Springer International Workshop on Information Hiding (IH)*, Darmstadt, Germany, 2009, pp. 235–246.
[43] R. Aggarwal, S. Singh, A. Kumar Roul, and N. Khanna, “Cellphone identification using noise estimates from recorded audio,” in *IEEE International Conference on Communications and Signal Processing (ICCSP)*, Melmaruvathur, India, 2014, pp. 1218–1222.
[44] M. Jahanirad, A. W. Abdul Wahab, N. B. Anuar, M. Y. Idna Idris, and M. N. Ayub, “Blind identification of source mobile devices using VoIP calls,” in *IEEE Region 10 Symposium*, Kuala Lumpur, Malaysia, 2014, pp. 486–491.

Electric Network Frequency Analysis

[45] C. Grigoras and J. M. Smith, “Advances in ENF analysis for digital media authentication,” in *AES International Conference on Audio Forensics*, Denver, CO, USA, 2012.
[46] ENFSI. (2009). “Best practice guidelines for ENF analysis in forensic authentication of digital evidence FSAAWG-BPM-ENF-001 (1.0).”
[47] C. Grigoras, “Digital audio recording analysis: The electric network frequency (ENF) criterion,” *The International Journal of Speech, Language and the Law*, vol. 12, no. 2, pp. 63–76, 2005.

[48] L. Cuccovillo and P. Aichroth, “Increasing the temporal resolution of ENF analysis via harmonic distortion,” in *AES International Conference on Audio Forensics*, Arlington, VA, USA, 2017.
[49] M. Fuentes, P. Zinemanas, P. Cancela, and J. A. Apolinário, “Detection of ENF discontinuities using PLL for audio authenticity,” in *IEEE Latin American Symposium on Circuits & Systems (LASCAS)*, Florianopolis, Brazil, 2016, pp. 79–82.
[50] D. P. Nicolalde Rodríguez, J. A. Apolinário, and L.W. Pereira Biscainho, “Audio authenticity: Detecting ENF discontinuity with high precision phase analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 534–543, 2010.
[51] M. Michałek, “The application of powerline hum in digital recording authenticity analysis,” *Problems of Forensic Sciences*, vol. 80, pp. 355–364, 2009.
[52] M. Huijbregtse and Z. Geradts, “Using the ENF criterion for determining the time of recording of short digital audio recordings,” in *Springer International Workshop on Computational Forensics (IWCF)*, The Hague, The Netherlands, 2009, pp. 116–124.
[53] C. Grigoras, “Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis,” *Forensic Science International*, vol. 167, no. 2-3, pp. 136–145, 2007.
[54] M. Kajstura, A. Trawínska, and J. Hebenstreit, “Application of the electrical network frequency (ENF) criterion: A case of a digital recording,” *Forensic Science International*, vol. 155, no. 2-3, pp. 165–171, 2005.
[55] N. Campos and A. Ferreira, “Real-time monitoring of ENF and THD quality parameters of the electrical grid in Portugal,” in *AES International Conference on Audio Forensics*, London, United Kingdom, 2014.
[56] A. Hajj-Ahmad, R. Garg, and M. Wu, “ENF-based region-of-recording identification for media signals,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 6, pp. 1125– 1136, 2015.
[57] Ž. Šarić, A. Žunić, T. Zrnić, M. Knežević, D. Despotović, and T. Delić, “Improving location of recording classification using electric network frequency (ENF) analysis,” in *IEEE International Symposium on Intelligent Systems and Informatics (SISY)*, Subotica, Serbia, 2016, pp. 51–56.
[58] J. Zjalic, C. Grigoras, and J. M. Smith, “A low cost, cloud based, portable, remote ENF system,” in *AES International Conference on Audio Forensics*, Arlington, VA, USA, 2017.
[59] C. Grigoras, J. M. Smith, and C. Jenkins, “Advances in ENF database configuration for forensic authentication of digital media,” in *131st AES Convention*, New York City, NY, USA, 2011.
[60] G. Hua, Y. Zhang, and V. L. L. Goh Jonathan; Thing, “Audio authentication by exploring the Absolute-Error-Map of ENF signals,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 1003-1016, 2016.
[61] C. Grigoras, “Applications of ENF analysis in forensic authentication of digital audio and video recordings,” *Journal of the Audio Engineering Society*, vol. 57, no. 9, pp. 643–661, 2009.

Copy-Move Forgery Detection

[62] M. Imran, Z. Ali, S. T. Bakhsh, and S. Akram, “Blind detection of copy-move forgery in digital audio forensics,” *IEEE Access*, vol. 5, pp. 12 843–12 855, 2017.
[63] M. Maksimović, L. Cuccovillo, and P. Aichroth, “Copy-move forgery detection and localization via partial audio matching,” in *AES International Conference on Audio Forensics*, Porto, Portugal, 2019.

Environment Analysis

[64] H. Malik and H. Farid, “Audio forensics from acoustic reverberation,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, 2010, pp. 1710–1713.
[65] H. Malik and H. Zhao, “Recording environment identification using acoustic reverberation,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, 2012, pp. 1833–1836.
[66] A. H. Moore, M. Brookes, and P. A. Naylor, “Room identification using roomprints,” in *AES International Conference on Audio Forensics*, London, United Kingdom, 2014.

Resampling Detection

[67] D. Vázquez-Padín and P. Comesaña, “ML estimation of the resampling factor,” in *IEEE International Workshop on Information Forensics and Security (WIFS)*, Costa Adeje, Spain, 2012, pp. 1833–1836.
[68] D. Vázquez-Padín, P. Comesaña, and F. Pérez-González, “Set-membership identification of resampled signals,” in *IEEE International Workshop on Information Forensics and Security (WIFS)*, Guangzhou, China, 2013, pp. 150–155.

Double-Encoding Detection

[69] T. Bianchi, A. De Rosa, M. Fontani, G. Rocciolo, and A. Piva, “Detection and classification of double compressed MP3 audio tracks,” in *ACM workshop on Information hiding and multimedia security*, Montpellier, France, 2013, pp. 159–164.
[70] Q. Huang, R. Wang, D. Yan, and J. Zhang, “AAC audio compression detection based on QMDCT coefficient,” in *Springer International Conference on Cloud Computing and Security (ICCCS)*, Haikou, China, 2018, pp. 347–359.
[71] R. Korycki, “Authenticity examination of compressed audio recordings using detection of multiple compression and encoders’ identification,” *Forensic Science International*, vol. 283, no. 1-3, pp. 54–67, 2014.

16 AMENDMENTS AGAINST PREVIOUS VERSION

Not applicable (first version).