www.czasopisma.pan.pl www.journals.pan.pl

ARCHIVES OF ACOUSTICS
Vol. **39**, No. 1, pp. 65–72 (2014)

DOI: 10.2478/aoa-2014-0007

Detection of Montage in Lossy Compressed Digital Audio Recordings

Rafał KORYCKI

Institute of Radioelectronics, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warszawa, Poland; e-mail: r.korycki@ire.pw.edu.pl

(*received January 21, 2013; accepted February 4, 2014*)

This paper addresses the problem of tampering detection and discusses methods used for authenticity analysis of digital audio recordings. Presented approach is based on frame offset measurement in audio files compressed and decoded by using perceptual audio coding algorithms which employ modified discrete cosine transform. The minimum values of total number of active MDCT coefficients occur for frame shifts equal to multiplications of applied window length. Any modification of audio file, including cutting off or pasting a part of audio recording causes a disturbance within this regularity. In this study the algorithm based on checking frame offset previously described in the literature is expanded by using each of four types of analysis windows commonly applied in the majority of MDCT based encoders. To enhance the robustness of the method additional histogram analysis is performed by detecting the presence of small value spectral components. Moreover, computation of maximum values of nonzero spectral coefficients is employed, which creates a gating function for the results obtained based on previous algorithm. This solution radically minimizes a number of false detections of forgeries. The influence of compression algorithms' parameters on detection of forgeries is presented by applying AAC and Ogg Vorbis encoders as examples. The effectiveness of tampering detection algorithms proposed in this paper is tested on a predefined music database and compared graphically using ROC-like curves.

Keywords: tampering detection, digital forgeries, digital audio authenticity, lossy compression, frame offsets, MDCT, AAC, Ogg Vorbis.

1. Introduction

Detection of forgeries in digital recordings is extremely important in juridical proceedings. Recordings may be useless as evidence when there is no proof that they are original and they have not been tampered. It is difficult to detect traces of montage in the era of digital audio files, due to the fact that currently available technologies e.g. free sound editing software allow a forger to change the meaning of uttered sentences without audible artifacts. Therefore, any tool that helps to evaluate digital audio authenticity may be of great importance to forensic audio experts. Nowadays, the most accurate and commonly used authentication method is based on electric network frequency (ENF) criterion (Grigoras, 2007). This approach utilizes random frequency fluctuations of the mains signal emitted by the electric network which are inadvertently induced in electronic circuits of recording devices. Therefore, its effectiveness is strictly dependent on the presence of mains signal in the recording, which occurs rarely. Furthermore some of the aspects of this approach, such as accurate measurements of ENF as well as searching and comparing the fluctuation patterns with a reference database are still being subjects of a scientific debate (Geradts, Huijbregtse, 2009; Korycki, 2010).

Recently, much attention is paid to authenticity analysis of compressed multimedia files. Several solutions were proposed for detection of double quantization in digital video which may result from multiple MPEG compression as well as from combining two videos of different qualities (Farid, Wang, 2009). Blocking periodicity analysis in JPEG compressed images was also investigated according to differences in quantization errors between neighboring blocks (Chen, Hsu, 2008). Due to differences between audio and picture compression algorithms it is impossible to adapt these methods in audio authenticity analysis. Instead, the detection of forgeries in compressed audio recordings must be based on other mathematical properties. Grigoras (2010) described statistical tools used to detect traces of audio recompression data to assess compression generation and also to discriminate between different audio compression algorithms. In the article (Liu *et al.*, 2010) authors presented novel approach to detect double MP3 compression by extracting the statistics of the modified discrete cosine transform (MDCT) spectral coefficients of MP3 signals, followed by applying a support vector machine. Huang *et al.* (2009) presented small-value MDCT coefficients as features to discriminate fake-quality MP3 from normal MP3 recordings.

Moreover, the number of active MDCT coefficients (NAC) was employed to perform authenticity analysis (*Huang et al.*, 2008). This approach can be successfully applied only to those recordings, in which the alleged forgery was made *after* audio files were decoded to lossless format, and without repeated encoding. The NAC minimums occur for frame shifts equal to multiplications of applied window length. Any modification of audio file, including cutting off or pasting a part of audio recording causes a disturbance within this regularity. Huang *et al.* (2008) presented an algorithm which detects forgeries in audio recordings compressed and decoded using MP3 encoder working with long blocks only. Although their method appears to be robust and accurate, the vast majority of compression algorithms use four types of analysis windows: long, short and two transition windows (start and stop). Therefore, based on the method described in (Huang *et al.*, 2008), three algorithms (ALG 1, ALG 2 and ALG 3) are introduced in this article, each of them with using four types of windows.

The ALG 1 consists in analysis of differences between adjacent minimum positions of NAC function. Depending on a window used during the encoding process, measured distances are matched with predefined pattern. Observed variances are compared for every window and stored as a possible indication of tampering. The ALG 2 algorithm employs additional histogram analysis. For each minimum of NAC function the absence of spectral components of magnitude values between 10^{-5} and 10^{-4} is examined. These two results are logically multiplied and the outcomes are treated in the same manner as minimums of NAC in ALG 1. In the last algorithm (ALG 3), detection of maximum values of NAC function is employed. The outcomes are then maximized within the area of window length creating a gating function for the results obtained based on ALG 2. The applied solution radically minimizes a number of false detections of forgeries.

The remainder of this paper is organized as follows. In Sec. 2 a brief introduction to lossy compression is given including short explanation of analysis windows and filter banks. MDCT properties, analysis of its coefficients and discussion on frame offsets is presented in Sec. 3. Shown in Sec. 4 are three tampering detection algorithms proposed by the author, which effectiveness in detection of forgeries will be proven based on database of music edited recordings. Obtained results are discussed in Sec. 5.

2. Lossy compression

Analyzing lossy compression algorithms, it is worth taking into consideration the principles that are relevant to the authentication process, i.e. spectral decomposition and quantization. The generic structure of perceptual audio coder consists of: (i) filterbanks and transform blocks where the input samples are converted into a subsampled spectral representation, (ii) perceptual model in which the signal's time-dependent masking threshold is estimated, (iii) quantization block where the quantization noise is masked, (iv) coding block, in which relevant information are packed into a bit stream (Herre, Schug, 2000).

In MPEG Layer 3 (MP3) encoder a sequence of 1152 input samples are polyphase filtered into 32 frequency subbands. The impulse responses of particular filters in this filter bank are defined as (Dabrowski, Marciniak, 2008):

$$h_i(n) = h(n)\cos\left[\frac{(2i+1)(n-16)\pi}{64}\right], \tag{1}$$

where $h(n)$ is an impulse response of the prototype lowpass filter. The output signal in i-th subband in the analysis filter bank is a convolution, as follows:

$$S_i(m) = \sum_{n=0}^{511} x(m-n)h_i(n). \tag{2}$$

Polyphase filter bank realization in the MP3 encoder is implemented, using a 512-point FIFO buffer. In each iteration 32 new input samples are inserted into the buffer and the all 512 samples are multiplied by the modified analysis window coefficients $C(n)$:

$$z(n) = c(n)x(n), \tag{3}$$

where

$$c(n) = \begin{cases} -h(n), & \text{if } n \text{ is odd,} \\ h(n), & \text{otherwise.} \end{cases} \tag{4}$$

Based on the Eqs. (1)–(4) the 32 new output samples are computed, as follows (Dabrowski, Marciniak, 2008):

$$S_j = \sum_{k=0}^{63} M_i(k) \sum_{j=0}^{7} z(k+64j), \qquad k = 0,1,\ldots,63; \qquad i = 0,1,\ldots,31, \tag{5}$$

where $M_i(k)$ are the modulation matrix coefficients:

$$M_i(k) = \cos\left[\frac{(2i+1)(k-16)\pi}{64}\right]. \tag{6}$$

Subsequently, Modified Discrete Cosine Transform (MDCT) is applied to the time frames of subband samples and each of the 32 subbands is split afresh into 18 subbands creating a granule with a total of 576 frequency coefficients. To reduce artifacts caused by time-limited operation on the signal, the windowing is applied. Depending on the degree of stationarity, the psychoacoustic model determines, which of four types of window is used (Huang *et al.*, 2008).

The Advanced Audio Coding (AAC) and Ogg Vorbis algorithms employ only MDCT filterbank which is switched between resolutions of 1024 and 128 spectral lines, depending on the stationary or transient character of the input signal. The conditions for a perfectly reconstructing oddly stacked time domain aliasing cancellation filter bank are derived in (Bradley, Princen, 1986). Assuming that analysis and synthesis filters have to be identical, followed equation must be satisfied:

$$w^2(k) + w^2(k+N) = 1, \tag{7}$$

where $w(k)$ is an analysis window having $2N$ samples. The sine window used in MP3 and AAC encoders is defined as (Malvar, 1990):

$$w(k) = \sin\left[\frac{\pi}{2N}\left(k+\frac{1}{2}\right)\right]. \tag{8}$$

By contrast, Ogg Vorbis algorithm employs quadratic sine window (Xiph.Org Foundation, Vorbis I specification, 2012, p. 11):

$$w(k) = \sin\left\{\frac{\pi}{2}\sin^2\left[\frac{\pi}{2N}\left(k+\frac{1}{2}\right)\right]\right\}. \tag{9}$$

Moreover, in AAC the shape of the transform window can be adaptively selected between a sine window and a Kaiser-Bessel-derived (KBD) window (Fulop, 2011):

$$w(k) = \frac{I_0\left(\pi\beta\sqrt{1-\left(\frac{2k}{N}-1\right)^2}\right)}{I_0(\pi\beta)}, \tag{10}$$

in which I_0 denotes the standard zero-order modified Bessel function. KBD window achieve better stop-band attenuation than the sine window, therefore for a pure tones more energy is concentrated into a single transform coefficient which reduces the perceptual bit allocation and improves coding gain. Nevertheless, the quantization noise is still inaudible since the uncoded coefficients have smaller magnitudes than the masked threshold (Atti *et al.*, 2007).

3. Analysis of MDCT spectral components

The MDCT is central for tampering detection algorithms, due to its properties. N-sized MDCT is computed based on $2N$ samples taken with 50% overlapping, according to Eq. (11):

$$X_{(p)}(k) = \frac{2}{N}\sum_{n=0}^{2N-1} x_{(p)}(n)\cdot w(n) \cdot \cos\left(\frac{\pi}{N}\cdot\left(n+\frac{N+1}{2}\right)\cdot(k+0.5)\right), \tag{11}$$

where $0 \le k \le N-1$. Applying Inverse-MDCT transform to the frame of spectral coefficients yields therefore $2N$ time-aliased samples:

$$\widehat{x}_{(p)}(n) = \frac{2}{N}\sum_{k=0}^{N-1} X_{(p)}(k) \cdot \cos\left(\frac{\pi}{N}\cdot\left(n+\frac{N+1}{2}\right)\cdot(k+0.5)\right), \tag{12}$$

where $0 \le n \le 2N-1$. These distortions are canceled by the overlap-and-add (OLA) procedure, which consists in computing an Inverse-MDCT, based on the previous and the next frame, multiplying each of the aliased segments by its corresponding window function and summing up overlapping time segments (Huang *et al.*, 2008):

$$x_(p)(n) = \begin{cases} \widehat{x}_{(p-1)}(n+N)\cdot w(N-n-1) + \widehat{x}_{(p)}(n)\cdot w(n), & 0 \le n \le N-1, \\ \widehat{x}_{(p)}(n)\cdot w(2N-n-1) + \widehat{x}_{(p+1)}(n-N)\cdot w(n-N), & N \le n \le 2N-1. \end{cases} \tag{13}$$

For signals with local symmetry, the MDCT coefficients are frequently reduced to zero (Huang *et al.*, 2008):

$$\begin{aligned} \widetilde{x}_p(n) &= \widetilde{x}_p(N-n-1), && 0 \le n \le N-1, \\ \widetilde{x}_p(n) &= -\widetilde{x}_p(3N-n-1), && N \le n \le 2N-1, \end{aligned} \tag{14}$$

where

$$\widetilde{x}_{(p)}(n) = x_{(p)}(n)\cdot w(n), \qquad 0 \le n \le 2N-1. \tag{15}$$

While encoding process is applied, spectral coefficients are quantized and some of them are assigned a zero value. This occurs for masked components as well as for unmasked parts, due to the probability distribution of the spectral coefficients and the compression ratio (Herre, Schug, 2000). Within the decoded signal, the troughs in a logarithmic spectral representation are visible only if identical framing offset to the one used for encoding is applied. It is essential to employ a correct decomposition algorithm, including proper window length and shape as well as the same filterbank type as used during the encoding process. This is because encoders' specifications usually do not define the exact steps for processing input data. The algorithms can therefore function quite differently and still satisfy the standard (Huang *et al.*, 2008).

Shown in Fig. 1 is MDCT analysis of one frame of decompressed signal which is shifted −1, 0, and +1 sample in respect to framing applied during ACC encoding process. As may be inferred from these examples, even a shift by one sample is sufficient to conceal the presence of characteristic zero values in the spectral representation of analyzed signal (Huang *et al.*, 2008). The troughs are visible only if the same frame offset is chosen as was employed during the encoding process.

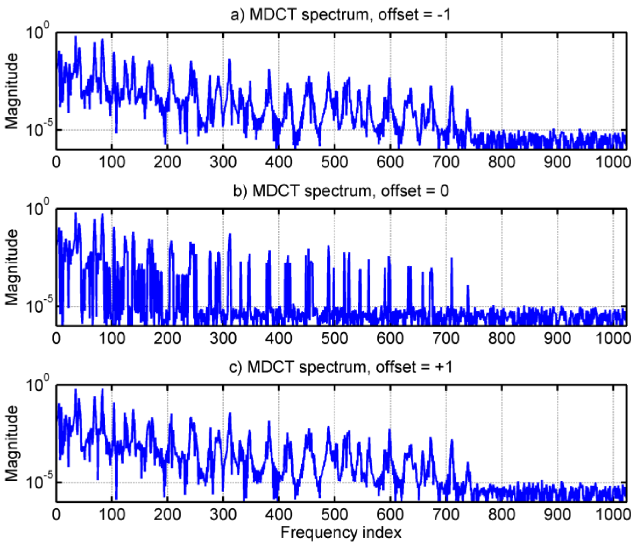

Fig. 1. MDCT coefficients of a decoded audio recording computed by using analysis window with: a) one sample left shift (offset = −1), b) no sample shift (offset = 0), c) one sample right shift (offset = +1) from the encoder frame grid, respectively. The magnitude is shown in the logarithmic scale.

The MDCT coefficients yield apparent significant distance, between the lowest peak and the highest valley of the spectrum for frame offset equal to zero. Shown in Fig. 2 are MDCT spectrum histograms of one frame of decompressed signal which is shifted −1, 0, and +1 sample in respect to framing applied during ACC encoding process. As can be seen, spectral components are not present between 10^{-5} and10^{-4} value, if the proper frame offset is chosen. It means that the magnitude of none of the spectral lines appears in this range of values. This phenomenon will be further utilized in automatic forgery detection algorithms.

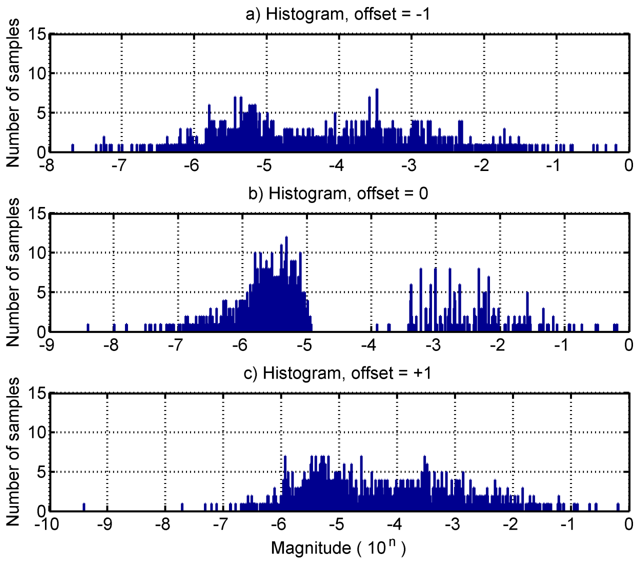

Fig. 2. Histogram of MDCT spectrum of a decoded audio recording computed by using analysis window with: a) one sample left shift (offset = −1), b) no sample shift (offset = 0), c) one sample right shift (offset = +1) from the encoder frame grid, respectively.

4. Tampering detection algorithms

The effects described in Sec. 3 are utilized in detection of forgeries in lossy compressed digital audio recordings. A number of active spectral coefficients (NAC) is calculated as proposed in (Huang *et al.*, 2008). For a given audio signal $\mathbf{x}$ of L samples long, a vector $\mathbf{x}^{(j)}$ is denoted as the vector after appending j zero samples at the beginning of $\mathbf{x}$, where $0 \le j < 1024$ for long and transition blocks and $0 \le j < 128$ for short blocks. For each shift j the vector $\mathbf{x}^{(j)}$ is split into 2048 (long and transition blocks) and 256 (short blocks) samples per frame with 50% overlapping, creating $N_L = (\lfloor L/1024 \rfloor - 1)$ frames for long and transition blocks and $N_S = (\lfloor L/128 \rfloor - 1)$ frames for short blocks:

$$\left[\widehat{x}_0^{(j)} \ldots \widehat{x}_{N-1}^{(j)}\right] = F\mathbf{x}^{(j)}, \tag{16}$$

where F represents the split operation as well as application of selected window function and $\widehat{x}_{(p)}^{(j)}$ is the p-th frame of $\mathbf{x}^{(j)}$. Thereafter, for each frame the MDCT is applied resulting in $X_{(p)}^{(j)}(k)$ which denotes the spectral representation of the p-th frame of $\mathbf{x}^{(j)}$.

Subsequently, the logarithm representation of the spectrum is calculated as proposed in (Huang *et al.*, 2008):

$$M_{(p)}^{(j)}(k) = \log_{10}\left(\max\left(X_{(p)}^{(j)}(k) \cdot X_{(p)}^{(j)}(k) \cdot 10^{10},\, 1\right)\right). \tag{17}$$

The number of active coefficients is then obtained by summing up all components of the logarithm representation of the spectrum obtained in (17):

$$C_{(p)}^{(j)} = \sum_{k=0}^{K-1} M_{(p)}^{(j)}(k), \tag{18}$$

where K represents the length of applied window: 1024 and 128, respectively. To locate the forgeries the offset is calculated for each of four types of windows as follows:

$$\text{offset}_{(p)} = \arg\min_{j} C_{(p)}^{(j)}. \tag{19}$$

If the calculated offset between adjacent minimums of $C_{(p)}^{(j)}$ differs, i.e.:

$$\text{offset}_{(p)} - \text{offset}_{(p-1)} \neq 0 \tag{20}$$

and no transition occurs (i.e. no minimum of $C_{(p)}^{(j)}$ for start or stop blocks), a forgery is indicated in frame p.

As stated above, computations are performed for each of four types of windows used by the encoder, which enables the algorithm to analyze real audio recordings. When current value of a frame shift equals multiplication of a window length, NAC reaches its minimum. Hence, if recording is tampered, the offset between adjacent minimums of NAC function outside the forgery position is other than current window length. Figure 3 illustrates a number of active spectral coefficients (NAC) related to the frame offset computed for each of four window types. An audio recording was compressed and decoded by AAC algorithm (with 128 kbps constant bit rate and 44.1 kHz sampling frequency), and afterwards, an editing process was performed by cutting out four fragments of the recording. To simplify the analysis, KBD window for long blocks (which is optional) was not applied in the AAC algorithm. Application of a short window requires a window switching sequence which is also apparent in Fig. 3. Another crucial observation is that edit points are accompanied by maximums of NAC

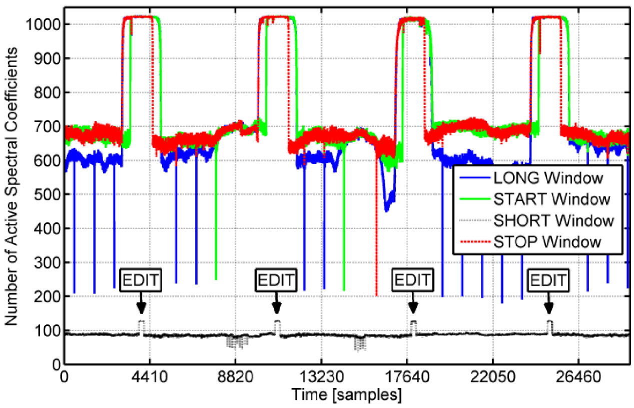

Fig. 3. NAC function related to the frame shift computed for a fragment of audio recording in which four edits were made. Audio recording was compressed and decoded using AAC algorithm with 128 kbps constant bit rate and 44.1 kHz sampling frequency.

function. Despite the fact that they may be found in other places where forgeries do not occur, these extreme values are utilized to improve robustness of tampering detection algorithm.

The NAC minimums occur only for frame shifts equal to multiplications of applied window length. Any modification of audio file, including cutting off or pasting a part of audio recording causes a disturbance within this regularity. Based on the aforementioned phenomenon, three algorithms were developed.

The ALG 1 consists in analysis of differences between adjacent minimum positions of NAC function, as described in (Huang *et al.*, 2008). Depending on a window used during the encoding process, measured offsets are matched with predefined pattern related to the window switching sequence. Starting with long blocks, the difference between adjacent offsets (see Eq. (19)) is calculated for each frame. If the Eq. (20) is satisfied, the distance between a minimum value of $C_{(p-1)}^{(j)}$ (for long blocks) and a minimum value of $C_{(p)}^{(j)}$ obtained for start blocks is calculated. Once the distance is equal to 1024, short blocks are being analyzed and the corresponding offset is being checked. Similarly, if the difference between adjacent offsets is not equal to zero for short blocks, the distance between a minimum value of $C_{(p-1)}^{(j)}$ (for short blocks) and a minimum value of $C_{(p)}^{(j)}$ (for stop blocks) is calculated. Thereafter, when this distance is equal to 576, long blocks are being analyzed afresh. Each time the Eq. (20) is satisfied and no transition occurs, a tampering is indicated.

The ALG 2 algorithm employs additional analysis in the ALG 1 when short blocks are being processed. The histogram count function shown in Eq. (21) is used to compute all components of magnitude values between 10^{-5} and 10^{-4} which are present in the MDCT spectra (see Fig. 2):

$$H_{(p)}^{(j)} = \text{histc}\Big(\log_{10}\Big(\Big|X_{(p)}^{(j)}(k)\Big|\Big),\ [-5,-4]\Big). \tag{21}$$

If the difference between adjacent offsets is not equal to zero, and $H_{(p)}^{(j)} > 0$, as well as no transition occurs, a tampering is indicated.

In the last algorithm (ALG 3), detection of maximum values of active spectral coefficients is employed (see Fig. 3). Each time the ALG 2 shows montage, the maximum value of NAC function is computed for the indicated frame and compared with a predefined threshold T:

$$\max C_{(p)}^{(j)} > T. \tag{22}$$

When the Eq. (22) is satisfied for the given frame, the ALG 3 indicates montage. The method employed in ALG 3 works similarly to the gating function, and therefore, radically minimizes a number of false detections of forgeries.

Theoretically, cutting out a part of audio file lasting the exact multiples of applied window might not have caused detectable disturbances in frame offset, therefore analyzed recording might be recognized as unaltered. However, application of MDCT computed for blocks of signal samples taken with 50% overlapping still allows to find the trace of this forgery. Shown in Fig. 4 is a number of active spectral coefficients (NAC) related to the frame offset computed for each of four window types. An audio recording was compressed and decoded by AAC algorithm (with 128 kbps constant bit rate and 44.1 kHz sampling frequency), and then a fragment of the recording lasting about 23 ms (10·1024 samples) was removed. As can be seen the distance between adjacent minimum positions of NAC function is other than 1024 for long blocks and the value of NAC function reaches its maximum in the vicinity of edit point.

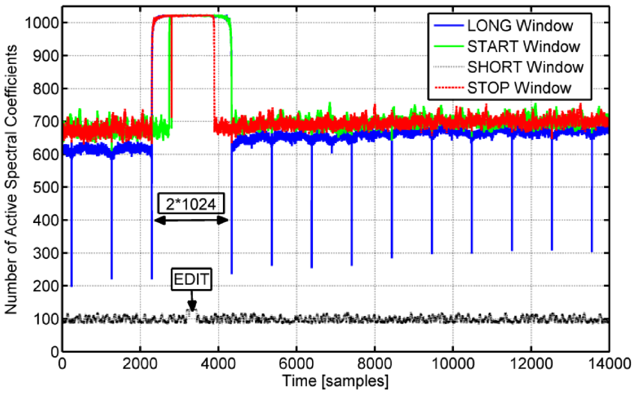

Fig. 4. NAC function related to the frame shift computed for a fragment of audio recording in which one edit was made. A part of audio file lasting the exact multiples of applied window length was removed.

5. Results and discussion

Described algorithms are examined for their usefulness as tools for detection of forgeries in lossy compressed audio recordings. The tests were conducted on a music database consisting of 15 music tracks with harmonic components and slowly changing audio background. The recordings were compressed and decoded using two different encoders with three different bit rates. In each of these tracks, 30 second long fragments were selected for further processing. The music database was prepared to mimic real forensic recordings, usually made in a noisy environment. In all of sampled fragments 21 deletions were performed at randomly selected locations and of randomly selected durations, however no longer than one second. Therefore, 315 forgeries were produced and subjected to further examinations.

The algorithms described in Sec. 4 were employed to detect edits performed in recordings from the music database. Thereafter, true acceptance ratios (TAR) and numbers of false acceptances (FA) were computed, given changing detection thresholds of minimums of NAC function as well as threshold value T denoted in Eq. (22). The TARs are obtained based on the number of detected edit points, divided by the known total number of forgeries. Shown in Table 1 and Table 2 are computational results obtained for two different encoders: AAC and Ogg Vorbis, respectively. Figures 5 and 6 show modified receiver operating characteristics (ROC) plotted for each of employed algorithms, encoders and bit rate values. Typically, ROC curve illustrates the performance of a binary classification system when its discrimination threshold is being modified. It is created by plotting the true positives out of the total number of positives as compared to the false positives out of the total number of negatives. The ROC-like curves used for the purpose of this article (Figs. 5, 6) depict true acceptance ratios in the function of the number of false acceptances, given changing detection threshold of minimums of NAC function.

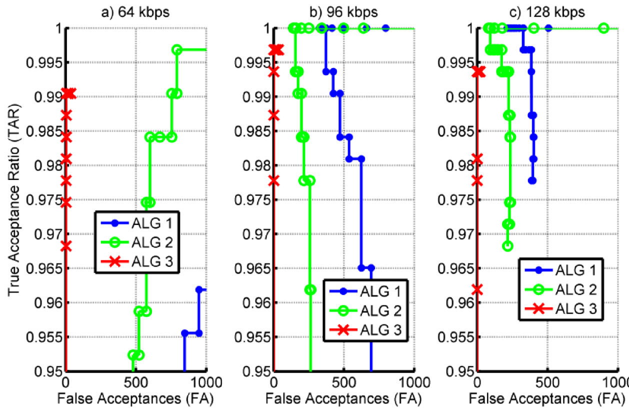

Fig. 5. ROC-like curves obtained during testing procedure of proposed algorithms executed on edited fragments from the music database. AAC compression algorithm was applied.

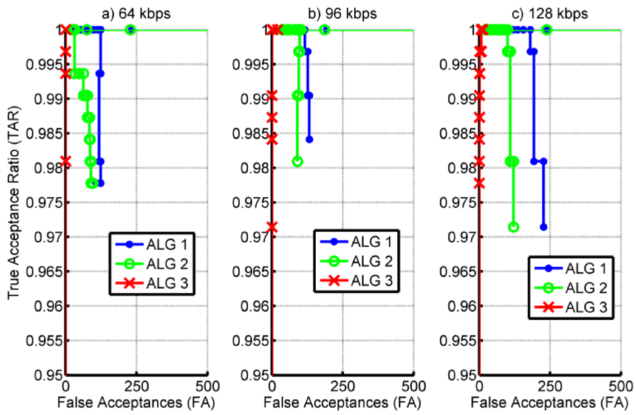

Fig. 6. ROC-like curves obtained during testing procedure of proposed algorithms executed on edited fragments from the music database. Ogg Vorbis compression algorithm was applied.

Table 1. Simulation results for edited fragments from the music database compressed and decoded using AAC encoder.

Bit rate [kbps]	ALG 1		ALG 2		ALG 3	
	TAR	FA	TAR	FA	TAR	FA
64	1.0000	1325	1.0000	1017	0.9905	6
96	1.0000	324	1.0000	140	0.9968	0
128	1.0000	214	1.0000	81	0.9810	0

Table 2. Simulation results for edited fragments from the music database compressed and decoded using Ogg Vorbis encoder.

Bit rate [kbps]	ALG 1		ALG 2		ALG 3	
	TAR	FA	TAR	FA	TAR	FA
64	1.0000	26	1.0000	21	1.0000	0
96	1.0000	48	1.0000	36	1.0000	0
128	1.0000	31	1.0000	22	0.9873	0

As may be seen, ALG 1 and ALG 2 algorithms, which are based on detecting minimums of NAC function give superior forgeries detection ratio with intolerable number of false acceptances. These results are coincident with the outcomes presented in (Huang *et al.*, 2008) and obtained for MP3 encoder using probably the long blocks only. In contrast, the ALG 3 algorithm yields slightly lower TAR values, while the number of false acceptances is reduced almost to zero. The results obtained for AAC encoder at 64 kbps (Fig. 5a) are significantly affected compared to all other bit rates, i.e. lower TAR values and higher number of false acceptances. This phenomenon is caused by default bandwidth limitation applied during encoding with the given bit rate.

6. Summary

The presented approach to analysis of nonzero spectral coefficients obtained based on MDCT transform can be successfully applied to those recordings in which the alleged forgery was made *after* audio files were decoded to lossless format and which were not encoded afresh. Described methods consisting in detection of frame offset of the compressed audio files can be successfully applied by forensic experts to detect forgeries in lossy compressed digital audio recordings. The value of tampering detection ratio for the given number of false acceptances equal to zero, is greater than 0.98. This allows the method to be recognized as a robust assistance in authenticity investigation process.

Notwithstanding its robustness and accuracy, presented approach requires more thorough research, especially in case of algorithmic differences between particular types of encoders. Future work will focus on evaluation and exploration of improved techniques for statistical analysis of the MDCT spectrum including machine learning algorithms. The indicated work will be employed to perform authenticity analysis, detection of double compression and identifying encoders' parameters. This research is currently being conducted by the author in the Forensic Laboratory of the Internal Security Agency of Poland within the project entitled "Designing empirical research and analyzing materials concerning specification of crime detection methods in public order special forces' operations", grant No. 0023/R/ID3/2012/02.

This article is a revised and extended version of a paper originally presented at the Joint Conference NTAV/SPA held in Łódź (Korycki, 2012).

Acknowledgments

The author acknowledges the financial support from the National Centre for Research and Development (Poland) in the form of grant No. 0023/R/ID3/2012/02.

References

1. Atti V., Painter T., Spanias A. (2007), *Audio Signal Processing and Coding*, John Wiley & Sons, Inc., Hoboken, New Jersey.
2. Bradley A.B., Princen J.P. (1986), *Analysis/synthesis filter bank design based on time domain aliasing cancellation*", IEEE Transactions on Acoustics, Speech and Signal Processing, **34**, 5, 1153–1161.
3. Chen Y., Hsu C. (2008), *Image tampering detection by blocking periodicity analysis in JPEG compressed images*, Proceedings of 10th Workshop on Multimedia Signal Processing, Oct. 8–10, pp. 803–808, Cairns, Queensland.
4. Dąbrowski A., Marciniak T. (2008), *Audio Signal Processing*, [in:] Digital Systems and Applications Oklobdzija V.G. [Ed.], Chapter 11, CRC Press, Davis, USA.
5. Farid H., Wang W. (2009), *Exposing Digital Forgeries in Video by Detecting Double Quantization*, Proceedings of ACM Multimedia and Security Workshop, Princeton.
6. Fulop S.A. (2011), *Speech Spectrum Analysis*, p. 32, Springer-Verlag, Berlin Heidelberg.
7. Geiger R., Herre J. Moehrs S. (2002), *Analysing decompressed audio with the Inverse Decoder – towards an operative algorithm*, Proceedings of 112th AES Convention, May 10–13, Munich, Germany.
8. Geradts Z., Huijbregtse M. (2009), *Using the ENF criterion for determining the time of recording of short digital audio recordings*, Proceedings of 3rd International Worshop on Computational Forensics, August 13–14, The Hague, The Netherlands.

9. Grigoras C. (2007), *Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis*, Forensic Science International, **167**, 136–145.

10. Grigoras C. (2010), *Statistical Tools for Multimedia Forensics: Compression Effects Analysis*, Proceedings of 39th AES International Conference Audio Forensics, June 17–19, Hillerod, Denmark.

11. Herre J., Schug M. (2000), *Analysis of Decompressed Audio – The Inverse Decoder*, Proceedings of 109th AES Convention, September 22–25, Los Angeles, USA.

12. Huang J., Qu Z., Yang R. (2008), *Detecting digital audio forgeries by checking frame offsets*, Proceedings of 10th ACM workshop on Multimedia and security – International Multimedia Conference, pp. 21–26, Oxford.

13. Huang J., Shi Y.Q., Yang R. (2009), *Defeating Fake-Quality MP3*, Proceedings of 11th ACM workshop on Multimedia and security – International Multimedia Conference, Sep. 7–8, Princeton NJ.

14. Korycki R. (2010), *Methods of Time-Frequency Analysis in Authentication of Digital Audio Recordings*, International Journal of Electronics and Telecommunications, **56**, 257–262.

15. Korycki R. (2011), *Research on Authentication of Compressed Audio Recordings*, Proceedings of 14th International Symposium on Sound Engineering and Tonmeistering ISSET 2011, May 19–21, Wroclaw, Poland.

16. Korycki R. (2012), *Detection of tampering in lossy compressed digital audio recordings*, Proceedings of Joint Conference New Trends in Audio and Video & Signal Processing: Algorithms, Architectures, Arrangements, and Applications NTAV/SPA 2012, September 27–29, Łódź, Poland, pp. 97–102.

17. Liu Q., Qiao M., Sung A.H. (2010), *Detection of double MP3 compression*, Cognitive Computing, **2**, 4, 291–296.

18. Malvar H.S. (1990), *Lapped transforms for efficient transform/subband coding*", IEEE Transactions on Acoustics, Speech and Signal Processing, **38**, 6, 969–978.

19. Xiph.Org Foundation, Vorbis I specification, Feb. 2, 2012, p. 11, http://www.xiph.org/vorbis/doc/Vorbis_I_spec.pdf