AMR Compressed-Domain Analysis for Multimedia Forensics Double Compression Detection

José Fabrizio Pereira Sampaio
Polícia Federal - Brasil

ABSTRACT

An audio recording must be authentic to be admitted as evidence in a criminal prosecution so that the speech is saved with maximum fidelity and interpretation mistakes are prevented. AMR (adaptive multi-rate) encoder is a worldwide standard for speech compression and for GSM mobile network transmission, including 3G and 4G. In addition, such encoder is an audio file format standard with extension AMR which uses the same compression algorithm. Due to its extensive usage in mobile networks and high availability in modern smartphones, AMR format has been found in audio authenticity cases for forgery searching. Such exams compound the multimedia forensics field which consists of, among other techniques, double compression detection, i. e., to determine if a given AMR file was decompressed and compressed again. AMR double compression detection is a complex engineering problem whose solution is still underway. In general terms, if an AMR file is double compressed, it is not compatible with an original one and may have been doctored. The published works in literature about double compression detection are based on decoded waveform of the AMR files to extract features. In this paper, a new approach is proposed to AMR double compression detection, which uses the encoded version to extract compressed-domain features based on linear prediction (LP) coefficients, in spite of processing decoded audio. By means of feature statistical analysis, it is possible to show that they can be used to achieve AMR double compression detection in an effective way. Therefore, compressed-domain features can be considered a promising path to solve AMR double compression problem by artificial neural networks.

Keywords: Multimedia Forensics. Audio Authenticity. AMR Encoder. Double Compression.

Recebido em 26 de março de 2018
Aceito em 15 de fevereiro de 2019

Brasília, v. 9, n. 2, p. 43-70, jul/dez 2018.

ISSN 2178-0013
ISSN Eletrônico 2318-6917

1. Introduction

Evidence analysis is a common procedure in criminal or civil court worldwide in which forensic examiners focus on answering prosecution or defense questions based on forensic science. Among all the possible fields, forensic science counts on multimedia forensics, which is a specialized field dedicated to exam multimedia evidence, like digital photography, digital audio, and video in order to adduce evidence to a legal process (BATTIATO, 2016). The first time multimedia forensics was recognized as a field of forensic science was in 2000. The exponential growth of multimedia evidence can be explained by the huge amount of stored digital information generated by, for instance, mobile phones and surveillance cameras. Specifically, audio forensics aims to exam audio evidence in many different ways, like to disclose forgery, to enhance poor intelligible speech, to elucidate a fact based on sound events (like gunshot analysis) or to identify speakers.

Audio forensics has a long story to tell, starting from the analog tapes (like in the Watergate iconic case) until facing digital audio. Due to the overwhelming predominance of digital techniques for storing audio, digital audio evidence has practically become the only audio evidence to be examined over the last fifteen years. This fact, jointly with the birth of many compression standards, promoted specialized research to create suitable techniques for audio authentication. A great review about audio authentication trends can be found in (GUPTA, 2012) and (ZAKARIAH, 2017).

Aiming to a literature review, for the sake of clarity, audio forensics can be divided into the following techniques, among others: electrical network frequency (ENF) criterion, frame offset, background noise and compression history identification. Before detailing them, however, one of the first techniques crated to detect digital audio forgeries should be cited. The bispectral analysis (FARID, 1999) uses polyspectral analysis and bicoherence to detect correlations introduced by nonlinear operations like a forgery, thus allowing examiners to find them. Nevertheless, such method is influenced by any nonlinearity during recording process, like compression, thus having a limitation for compressed digital audio.

The ENF criterion is a well-established technique to authenticate digital audio becasue it uses the slow frequency and phase variations of electrical network power signal as a reliable reference to check continuity, date of recording and to pinpoint splices (GRIGORAS, 2003). Among many published works upon ENF technique, some may be picked to overview the method. Whenever a reliable ENF database exists and the audio specimen presents an ENF tone, it is possible to determine date of recording and exact audio location of forgery, even if the recorder is battery powered. However, the ENF detection is a difficult task if this signal is too weak (KAJSTURA, 2005). The use of ENF criterion is feasible even without a database if the signal phase is inspected by means of a high precision DFT (discrete Fourier transform) to detect phase discontinuity and disclose tampering (RODRIGUEZ, 2010). In addition, the ENF variation tracking is an interesting approach to detect signal discontinuity because unauthentic audio presents anomalous ENF variations that can be detected with a variable threshold (ESQUEF, 2014) or with an outlier detector based on the samples kurtosis via SVM (support vector machine) classifier (REIS, 2017). Another ENF application is the identification of recaptured audio because the ENF tone is double imprinted whenever the original audio is recaptured somehow, like in analog to digital conversion or in recording while playing in different recorders, which permits to identify a forgery or an anti-forensic strategy (SU, 2013). Although quite a few works using ENF were mentioned in this paper, among many already published, ENF criterion is actually recognized as the most reliable forensic method to detect audio forgeries.

The frame offset methods rely on peculiarity of encoder algorithm. For MP3 encoder, which is the most popular audio format up to date, the frame offset technique is the most reliable tool to pinpoint common tampering like deletion, insertion, substitution and splicing. By using the number of active MDCT (modified discrete cosine transform) coefficients, the frame grid can be verified for forgeries as the frame structure is broken after alterations. Such method permits to locate the position the forgery was made in the audio file with more than 99% detection rate, even in low bitrates (YANG; QU, 2012). However, when MP3 file is transcoded or the bitrate is too high, the method does not perform well. An enhanced method is considered by using different analysis windows and additional histogram analysis to increase robustness and detection rates (KORYCKI, 2014).

Regarding forensic analysis of audio background noise, the room noise and acoustic reverberation profiles, as well as recorder noise, have been a useful tool to detect audio forgeries. Background noise should be extracted with minimal speech content to be used in audio authentication to compare different environments and noise pattern estimates, considering that an original recording contains the same noise pattern or smooth transitions (IKRAM, 2010). Acoustic reverberation is a valuable measurement to classify environment type based only on recorded audio. The statistical technique to estimate the amount of reverberation variance, associated with an automatic decaying-tail-selection, permits to identify recording environments acoustically (MALIK, 2013). Audio background noise and speech itself are also the basis for device identification, which is an important issue in multimedia forensics. The identification of eight telephone handsets and microphones is possible by means of speech mel frequency cepstral coefficients (MFCC) parameterization and Gaussian supervectors, achieving a 90% classification accuracy via SVM (ROMERO, 2010). Better results were reported about the task of identifying mobile phone microphones using non-speech segments of speech recordings (i.e., background noise), reaching identification rates of 96.42% (HANILÇI, 2014).

Another field of multimedia forensics consists of compression history identification, which has caught research attention in the past eight years. The basic idea is to find out the compression algorithms used to process a given audio specimen, even if such audio is in uncompressed format, like linear pulse code modulation (PCM) WAV, because previous compression may leave detectable traces and reveal forgeries. Another case study is to determine whether an audio file has been compressed twice by the same compression algorithm, which is called double compression. One of the first reported studies upon MP3 double compression detection used first digit probability distributions of MDCT coefficients because double compressed MP3 audio violates Benford's Law (YANG, 2010), thus permitting to extract features to SVM classification. One practical case is related to fake-quality MP3 files sold on the Internet as original ones, that is, the label bitrate is different from real bitrate. One method to detect MP3 low to high bitrate transcoding, which is double compression itself, is to inspect quantized MDCT coefficient values, because transcoded audio has fewer MDCT coefficients of small values (YANG, 2009). Other methods use features based on average number of zero dequantized MDCT coefficients (QIAO, 2010) or based on statistics of MDCT coefficients (LIU, 2010), all using SVM for method accuracy evaluation. Besides MP3 compression history identification, there are works concerning other encoders, like AAC, AMR, WMA and uncompressed WAV. The audio codec identification without decoding to time domain is possible by considering chaotic characteristics of bitstream, even if the audio is first compressed with another encoder (HIÇSÖNMEZ, 2011). Compression history can be identified in detail, including first and second compression codecs and bitrates, as well as single and double compressed audio can be discriminated by using audio quality measures or statistics over encoded bitstream (HIÇSÖNMEZ, 2013). If the considered audio file is already decompressed, like in WAV format, the previous compression operations can be disclosed, including MP3 or WMA codec usage and respective bitrates, by means of a set of MDCT features (LUO, 2012). The identification becomes more accurate (near 100% for some encoders) when a profile of a set of noise spectra and a time-domain histogram are used (JENNER, 2012). More recently, a MDCT and MFCC feature-based method achieved accuracies about 99% to identify MP3, WMA and AAC previous compression bitrates in WAV files (LUO; LUO, 2014), including high bitrates like 128 kbits/s, which is a complicated task due to high similarity to uncompressed PCM audio.

AMR double compression analysis is a specific case of the compression history issue, but exploring a different encoder from MP3 and admitting recompression using the same encoder. Instead of considering MDCT properties or frame offset analysis, existing AMR double compression detection methods focus on machine learning algorithms to solve this complex problem, as AMR algorithm does not rely on any transform. The first method found to detect AMR double compression uses acoustic features extracted from decoded audio and SVM for discrimination, reaching average accuracy about 87% (SHEN, 2012). Two years later, a deep learning-based method was reported with average accuracy about 91%. Such method uses raw decompressed AMR audio as the input to a three layer deep learning classifier (LUO; YANG, 2014). More recently, a stacked autoencoder-based method achieved better accuracies about 98% for AMR double compression detection, once again using raw decompressed AMR as the input. The main idea of such method is the automatic feature extraction by stacked autoencoder neural network, which significantly increases accuracies. In addition, the robust analysis against frame offset attack, noise, different corpora, and variable duration audio showed this method is reliable (LUO, 2017).

As the literature review shows, AMR double compression detection is an interesting multimedia forensics topic with relevant published works. In this piece of work, a new approach is proposed to build features used as input to a neural network and possibly achieve better performance than state-of-the-art. Instead of using decoded AMR audio to extract features, compressed-domain features are proposed based on linear prediction (LP) coefficients. Such idea of extracting compressed-domain features, although being a novel approach for AMR double compression detection, is a well known topic in signal processing community (PFEIFFER, 2003), like in automatic speaker recognition for AMR audio in compressed domain (PETRACCA, 2005).

This article is divided in three sections besides this introduction. In Section 1, a review on AMR codec is provided, including the double compression procedure during the tampering process. Section 2 gives specific details about LP coefficient extraction from encoded AMR audio. The experimental setup is described in Section 3, including statistical computation and analysis of LP coefficients extracted from single and double compressed AMR files. In conclusion, the possibility of using the statistics of LP coefficients as features is addressed, gathering examples of neural networks to be used. The future works are also stated in conclusion.

2. AMR Encoder/Decoder and Double Compression Inevitability

The AMR standard (3GPP, 2011) is present in every GSM mobile phone call to convey speech to and from a mobile phone, aside from defining the AMR digital audio file format with constant bitrate (extension AMR) and some smartphones voice notes applications. Such encoder/decoder (codec hereinafter) was engineered to transmit speech over the mobile radio channel and compensate for channel fading by means of an adaptive bitrate. The sampling rate is 8,000 Hz and there are 8 possible bitrates as follows (rates in kbits/s): 12.2, 10.2, 7.95, 7.4, 6.7, 5.9, 5.15 and 4.75.

The basis of AMR codec is the CELP (code excited linear prediction) algorithm which encodes voiced frames (like vowels) using an LP filter and non voiced frames (like consonants) using adaptive codebooks. AMR uses an LP filter H(z) with 10 (m=10) LP coefficients ai according to the definition:

$$H(z) = \frac{1}{1 + \sum_{i=1}^{m} a_i z^{-i}} \qquad (1)$$

The LP coefficients represent an all-pole linear filter derived from the source-filter model of speech production. This filter is a model of the vocal tract of the speaker and the LP coefficients define the frequencies of the speaker formants.

Starting from sampled speech, AMR encoder processes 160-sample frames one at a time to compute parameters like LP coefficients, pitch lags and codebooks gains and indexes. Each encoded frame is represented by an AMR frame to form the bitstream, which content is the quantized representation of parameters with bit allocation depending on AMR bitrate. For instance, LP coefficients are not directly present in AMR bitstream because a vector quantization and a nonlinear transformation are desirable to make LP coefficients less prone to transmission errors over the radio channel.

Whenever a forger wants to tamper with a compressed audio file, like making suppression or splices, he/she has to decode the original file to time domain, make the forgery, and compress again to a compressed format which is normally the same as the original file to keep recorder compatibility, characterizing the double compression. The AMR file frame structure offers encoded frames without direct access to time domain samples, becasue the bitstream only offers the quantized AMR parameters. In other words, it is unfeasible to manipulate the speech directly in encoded audio and, therefore, a decoding is necessary. If the final file format is supposed to be the same as original, the double compression takes place as an inevitable procedure to forgers.

3. Linear Prediction Coefficient Extraction

As LP coefficients are not directly present in AMR bitstream, an effective way to extract them from encoder algorithm is to analyze AMR decoder source code. This method can provide the LP coefficient values, i.e., before transformation and quantization, which are the coefficients of the LP filter. By definition, LP coefficients have the spectral meaning of defining the vocal filter, that is to say, the shape of speaker vocal tube and formant locations. This meaning leads to the interpretation that if double compression distorts speech somehow, as it actually appears, the LP coefficients may change, at least in terms of their statistics. That is why extracting LP coefficients may be useful to analyze double compression.

AMR source code analysis, written in C language, permits to extract LP coefficients in binary files by inspecting the variable A_t[ ] and modifying files decoder.c and sp_dec.c. After recompiling AMR source code, a new version of decoder will decode AMR files and generate the LP coefficient binary files. The output binary files contain 40 LP coefficients for each frame because the LP filter has 10 coefficients and each frame is divided in four subframes.

Considering an AMR file to analyze, instead of decoding it to time domain, it is partially decoded to extract LP coefficients in number proportional to speech duration in file. To be precise, the number of LP coefficients extracted will be as large as the number of voiced frames detected in speech.

4. LP Coefficient Statistical Behavior in Double Compression

Bearing in mind that a given AMR file generates a chain of speech dependent LP coefficients, direct comparison one by one of such coefficients between single and double compressed files would probably not give acceptable results. As the literature shows, the AMR double compression detection is evaluated by comparing one single compressed file with its own double compressed version at the same or different first bitrate, i.e., the speech content is the same and the first and second bitrates may be different (SHEN, 2012). Therefore, one approach is to analyze LP coefficient behavior by choosing some statistical averages to reveal possible differences between single and double compressed files, which will be represented by such averages to be compared.

4.1 Experimental Setup

An audio corpus (a set of utterances recorded in the same conditions) must be considered with enough samples to build reliable statistics. The TIMIT database (GAROFOLO, 1993) is an appropriate speech corpus used in many works in multimedia forensics, like in (ROMERO, 2010), (HANILÇI, 2014), (JENNER, 2012), (SHEN, 2012), (LUO; YANG, 2014) and (LUO, 2017). Such database has 6,300 uncompressed files of variable durations (about 1 second, but some as long as 7 seconds) recorded in the same conditions by 630 different United States of America speakers saying 10 sentences.

By using the original AMR codec, 6,300 single compressed files at 4.75 kbits/s were generated. These single compressed files were decoded and encoded again at 4.75 kbits/s to generate 6,300 double compressed files. After doing so, the modified AMR decoder described in Section 2 was used to partly decode all the 12,600 files, generating 6,300 binary files with the respective single compressed LP coefficients and 6,300 binary files with the respective double compressed LP coefficients.

4.2 STATISTICAL ANALYSIS

Each LP coefficient binary file should be processed to compute statistical figures to be compared. Table 1 shows the 6 chosen statistical averages to compare each AMR file. As it can be seen, such averages are simple and they only intend to enlighten any existing difference between single and double AMR compressed audio.

Definition	**Symbol**
Mean	$\overline{x}$
Standard Deviation	σ
Mode	*Mo*
Median	$\widetilde{x}$
Mean of Squared Elements	*meansqr*
Median Absolute Deviation	MAD

Table 1. Statistical averages chosen to analyze LP coefficient behavior in single and double AMR compression

The statistical analysis was made by using three different bi-dimensional graphic representations: histograms, scatter plots, and boxplots.

The histogram is a plot of data distribution which is counted (Y axis) after fitting values in bins (X axis). In this work, the bins are all the same size and correspond to ranges of the averages in Table 1. As there are 6,300 LP coefficient binary files for single and double compressed files, there will be, for instance, 6,300 mean values for single compressed (and also for double compressed) and the histogram of all the mean values for single compressed (or double) will count 6,300 for all the bins. The plots in this work have always two histograms: the black bars for single compressed files and the white bars for double compressed files. For overlapped histograms the bars are all in gray.

The scatter plot is a direct comparison of average values for single and double compressed files. The X axis corresponds to one of the averages of Table 1 for the single compressed files and the Y axis corresponds to the respective averages for double compressed files. In other words, a dot in the scatter plot is defined by the single compressed average in X axis and the double compressed average in Y axis for the same file after double compression. The identity line (flagged as a red line) is an important point of observation, as this line can cross the scatter in the middle (no evident tendency after double compression), below (evident rising tendency after double compression) or above (evident decrease tendency after double compression).

The boxplot is a well known representation of data distribution. It consists of a squared box where the median is showed as a line crossing the box and the quartiles and outliers (red crosses) are also evident. In this work, the figures have always two boxplots, one for 6,300 single compressed files (on the left) and another for 6,300 double compressed files (on the right).

As the statistical averages can be computed for each LP coefficient independently (1st until 10th which corresponds to a1 until a10), there are 60 possible measurements for a given file. In terms of machine learning terminology, each file can be represented by a 60-element feature vector, that is to say, a computed average is defined as a feature. For the sake of objectivity, some features of chosen LP coefficients were selected as best and fair examples of discrimination capability between single and double compressed AMR files. The first compression bitrate and the second compression bitrate are both 4.75 kbits/s to all the following figures.

Figures from 1 to 3 show the mean of the 7th LP coefficient. The histograms show evident statistical differences as the double compression occurs. The scatter plot shows a global decrease in values and the boxplots confirm the differences. The mean of the 7th LP coefficient can therefore be considered a good feature do detect AMR double compression.

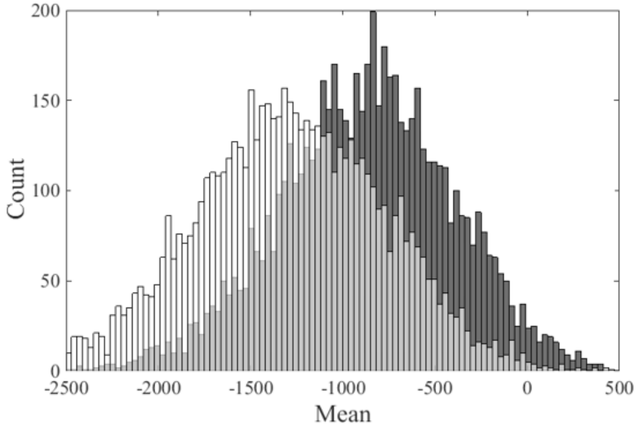

Figure 01. Histogram of mean of 7th LP coefficient. Note the distribution left displacement after double compression.

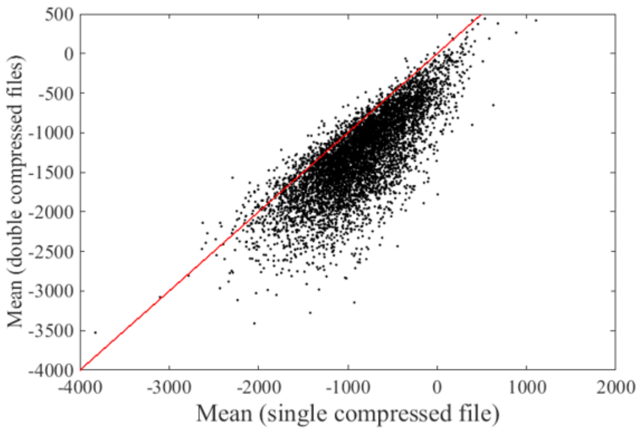

Figure 02. Scatter plot of mean of 7th LP coefficient. Note the decrease in mean values after double compression as the scatter average position is below the identity line.

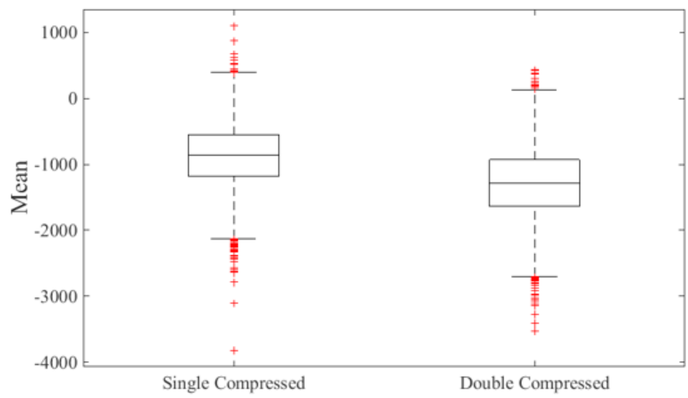

Figure 03. Boxplot of mean of 7th LP coefficient. Note the median, extremes and quartiles decrease after double compression.

Figures from 4 to 6 show the standard deviation of the 8th LP coefficient. The histograms show evident statistical differences as the double compression occurs. The scatter plot shows a global increase in values and the boxplots confirm the tendency. The standard deviation of the 8th LP can therefore be considered a good feature to detect AMR double compression.

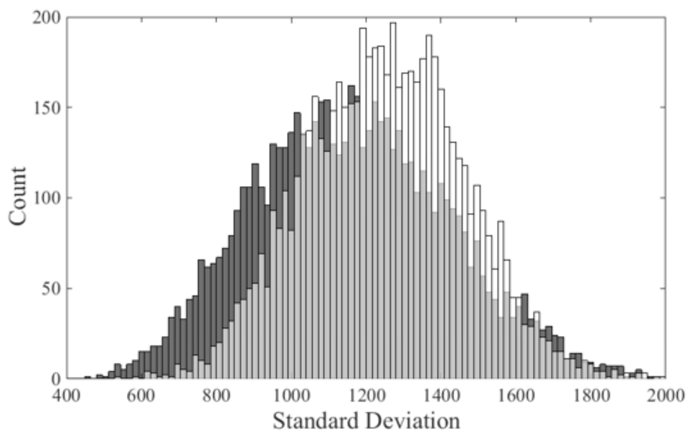

Figure 04. Histogram of standard deviation of 8th LP coefficient. Note the distribution right displacement after double compression.

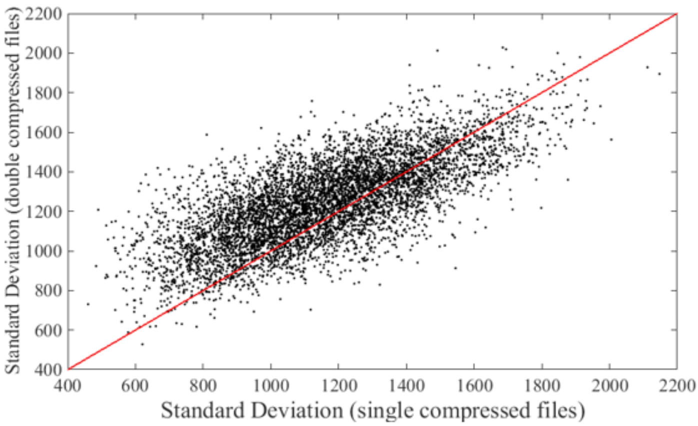

Figure 05. Scatter plot of standard deviation of 8th LP coefficient. Note the increase in standard deviation values after double compression as the scatter average position is above the identity line.

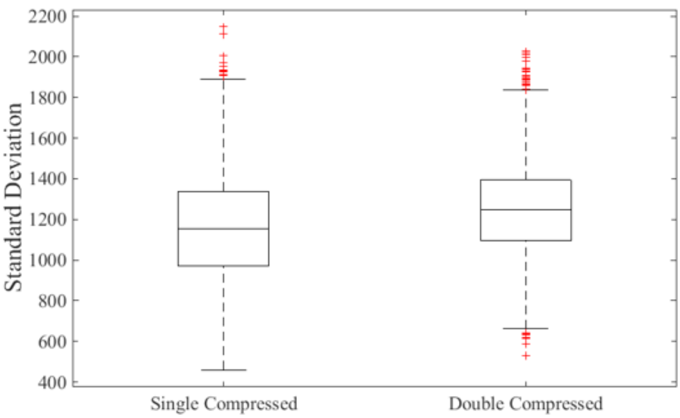

Figure 06. Boxplot of standard deviation of 8th LP coefficient. Note the median, extremes and quartiles increase after double compression.

Figures from 7 to 9 show the mode of the 8th LP coefficient. The histograms show slight statistical differences as the double compression occurs. The scatter plot shows a slight global increase in values and the boxplots confirm the tendency. The mode of the 8th LP can therefore be considered a fair feature to detect AMR double compression.

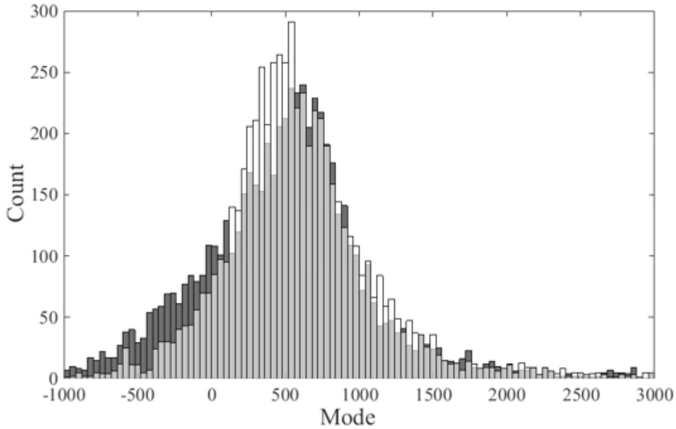

Figure 07. Histogram of mode of 8th LP coefficient. Note the slight distribution mode left displacement after double compression.

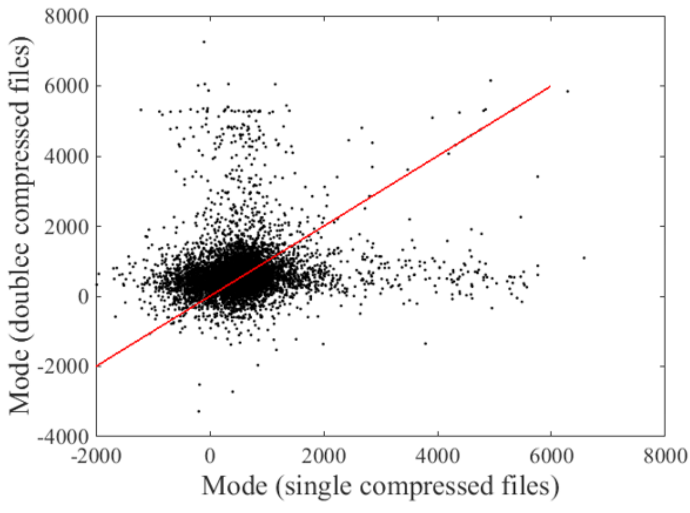

Figure 08. Scatter plot of mode of 8th LP coefficient. Note the scatter is slightly above the identity line.

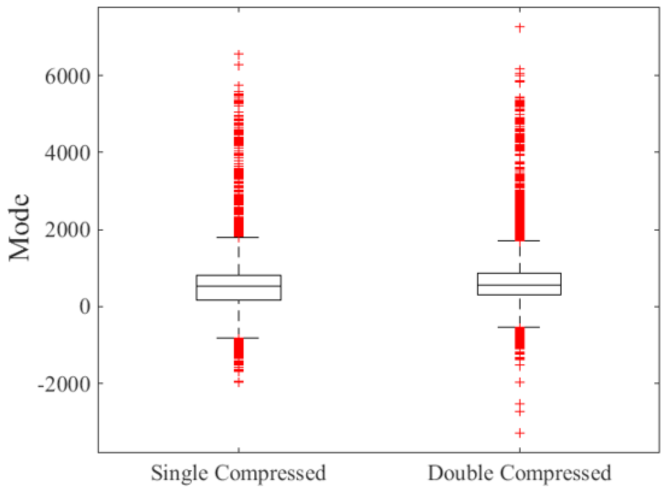

Figure 09. Boxplot of mode of 8th LP coefficient. Note the slight median, extremes and quartiles variations after double compression.

Figures from 10 to 12 show the median of the 9th LP coefficient. The histograms show evident statistical differences as the double compression occurs. The scatter plot shows a global decrease in values and the boxplots confirm the tendency. The median of the 8th LP can therefore be considered a good feature to detect AMR double compression.

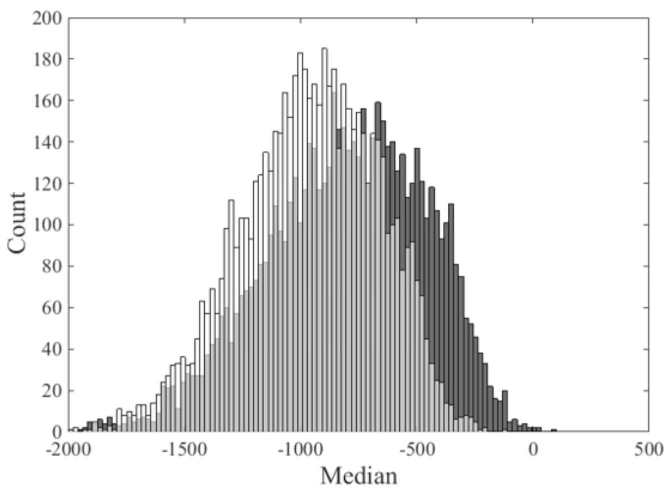

Figure 10. Histogram of median of 9th LP coefficient. Note the distribution mode left displacement after double compression.

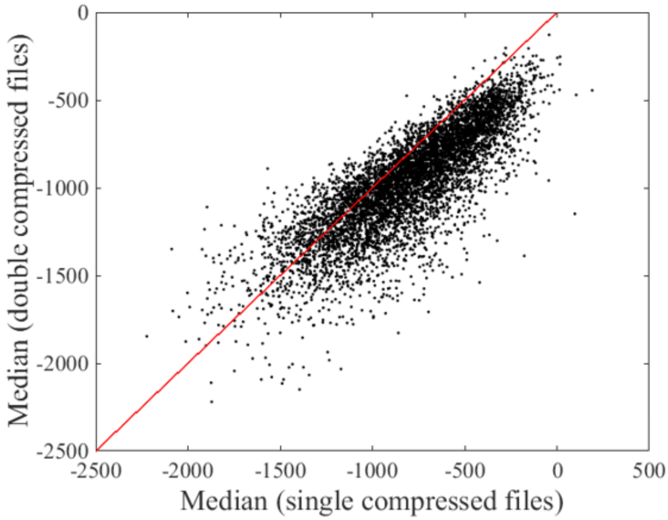

Figure 11. Scatter plot of median of 9th LP coefficient. Note the decrease in median values after double compression as the scatter average position is under the identity line.

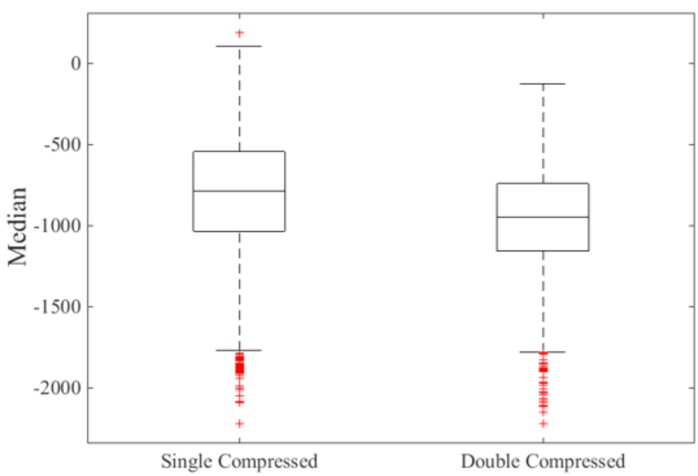

Figure 12. Boxplot of median of 9th LP coefficient. Note the median, upper extreme and quartiles decrease after double compression.

Figures from 13 to 15 show the mean of squared elements of the 4th LP coefficient. The histograms show evident statistical differences as the double compression occurs. The scatter plot shows a global increase in values and the boxplots confirm the tendency. The mean of squared elements of the 4th LP can therefore be considered a good feature to detect AMR double compression.

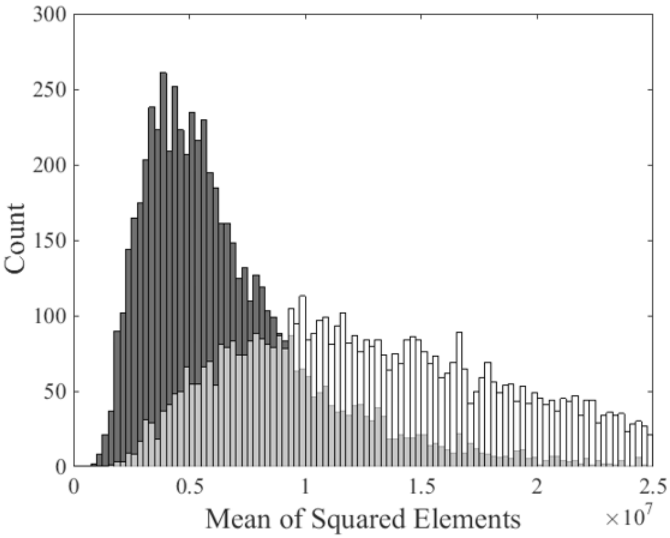

Figure 13. Histogram of mean of squared elements of 4th LP coefficient Note the distribution right displacement and broadening after double compression.

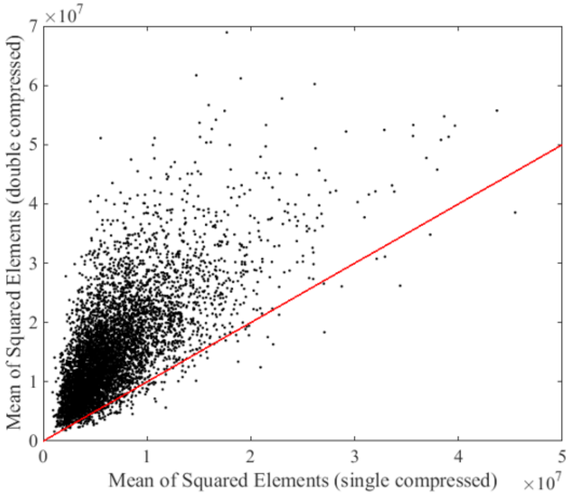

Figure 14. Scatter plot of mean of squared elements of 4th LP coefficient. Note the increase in mean of squared elements values after double compression as the scatter average position is above the identity line.

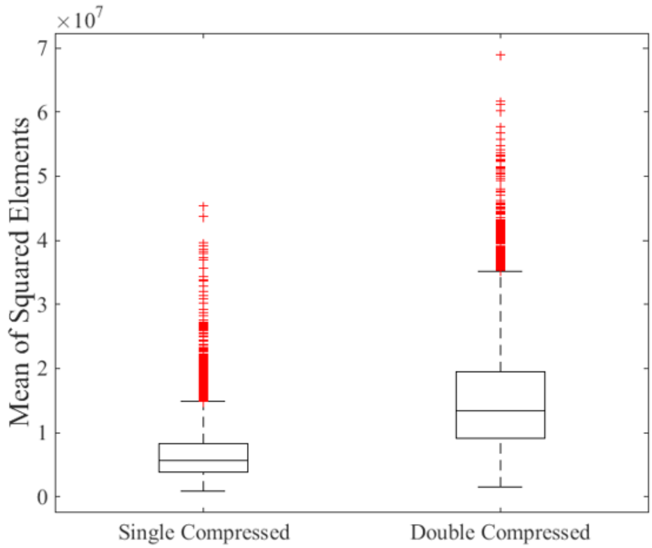

Figure 15. Boxplot of mean of squared elements of 4th LP coefficient. Note the median, extremes and quartiles increase after double compression.

Figures from 16 to 18 show the median absolute deviation of the 6th LP coefficient. The histograms show evident statistical differences as the double compression occurs. The scatter plot shows a global increase in values and the boxplots confirm the tendency. The median absolute deviation of the 6th LP can therefore be considered a good feature to detect AMR double compression.

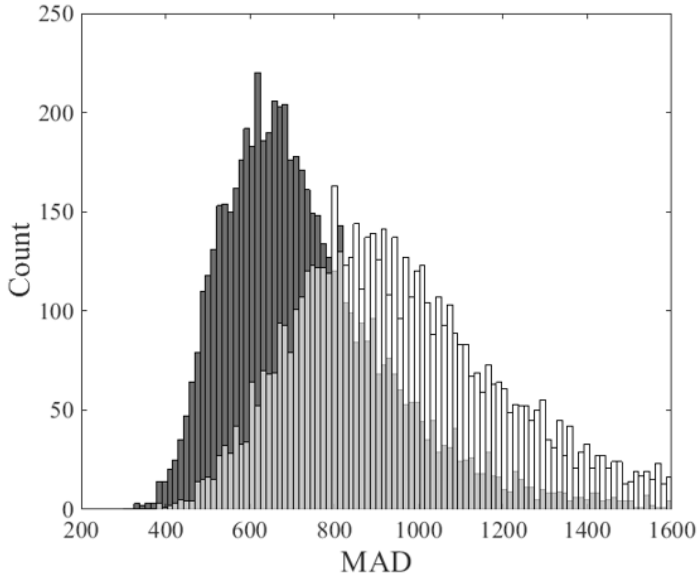

Figure 16. Histogram of median absolute deviation of 6th LP coefficient. Note the distribution right displacement and broadening after double compression.

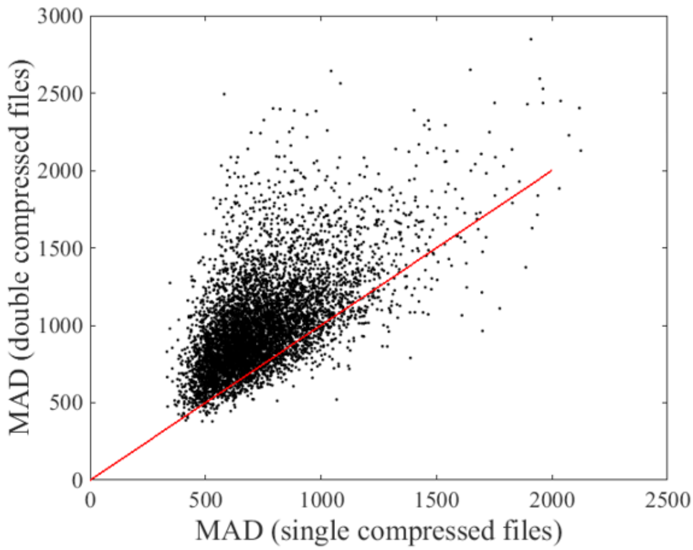

Figure 17. Scatter plot of median absolute deviation of 6th LP coefficient. Note the increase in median absolute deviation values after double compression as the scatter average position is above the identity line.

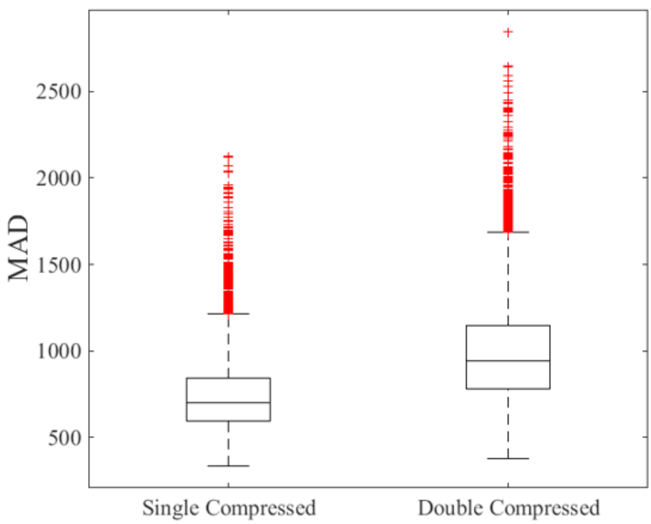

Figure 18. Boxplot of median absolute deviation of 6th LP coefficient. Note the median, extremes and quartiles increase after double compression.

5. Conclusion and Future Work

Multimedia forensics is a relative new area of forensic science dedicated to process multimedia evidence, like audio, images, and video. AMR digital audio must be authentic to be admitted as evidence in court cases. A useful technique to assess AMR authenticity is to determine if it is single or double compressed, but such a question consists of a complex and ongoing-solution engineering problem. The state-of-the-art and previous works for AMR double compression detection uses only decompressed audio. In this paper, a new approach is proposed to detect AMR double compression by means of compressed-domain averages based on linear prediction coefficients.

The statistical analysis shows that simple LP coefficient averages can be used as features of a given AMR file because the double compression procedure disturbs the statistics in a discriminative tendency. The experiments with TIMIT database show that some features perform better than others in discrimination, but all of them present consistent modification after double compression.

The presented features are strong candidates to be used all together as inputs to a neural network, like support vector machines (SVM), self organizing maps (Kohonen maps) or deep learning networks. Such neural networks have the ability to consider all the feature behavior simultaneously and predict after supervised or unsupervised learning.

For future work, a greater number of features will be considered as the AMR double compression detection complexity may ask for more measurements. Additional statistical features may be used as well as LP coefficient first digit probabilities (similar to MDCT coefficients in MP3 double compression detection). Another possible modification is to compute statistics for other AMR encoder parameter, like line spectral pairs (LSP) and pitch lags. After increasing the number of features, experiments using an appropriate neural network will give numeric results of the chosen features discrimination capability and may indicate if the proposed approach may surpass the state-of-the-art results.

José Fabrizio Pereira Sampaio
Perito Criminal Federal
Engenheiro eltrônico pelo Instituto Militar de Engenharia (IME/RJ)
Mestre em Ciências pela Universidade de Brasília
Doutorando pela Universidade de Brasília
E-mail: fabrizio.jfps@dpf.gov.br

Análise Forense de AMR Compressed-Domain para Detecção de Dupla Compressão Multimídia

RESUMO

Para ser aceita como prova no processo criminal, uma gravação de áudio deve ser autêntica de modo a se preservar com a máxima fidelidade possível as falas registradas e evitar erros de interpretação. O codificador AMR é um padrão mundial para compressão e transmissão de voz nas redes de telefonia móvel GSM 3G e 4G. Além disso, tal codificador é um padrão de armazenamento de arquivos de extensão AMR que usa o mesmo algoritmo de compressão. Pelo seu extenso uso nas redes móveis e alta disponibilidade nos modernos smartphones, o formato AMR tem sido encontrado em casos de exames de autenticidade de áudio em que se questiona se houve alguma manipulação. Tais exames fazem parte da área de multimídia forense que é formada, dentre outras técnicas, pela detecção de dupla compressão, ou seja, determinar se um arquivo AMR foi descomprimido e comprimido novamente. A detecção da dupla compressão AMR é um problema complexo de engenharia cuja solução ainda está em curso. Em linhas gerais, se um arquivo AMR tem dupla compressão, ele não é compatível com um original e pode ter sido modificado. Os trabalhos publicados na literatura sobre detecção de dupla compressão utilizam o arquivo AMR decodificado no domínio do tempo como ponto de partida para a extração de características. Neste artigo, é proposta uma nova abordagem para o problema da detecção de dupla compressão AMR em que, ao invés de se utilizar o áudio decodificado, é usada a sua versão codificada para a extração de características no domínio da compressão baseadas nos coeficientes de predição linear (LP). Por meio de análise estatística dessas características, é possível mostrar que elas podem ser usadas para realizar a detecção de dupla compressão AMR de forma efetiva. Dessa forma, as características no domínio da compressão podem ser consideradas um caminho promissor para a resolução do problema da dupla compressão AMR por meio de redes neurais artificiais.

Palavras-Chave: Multimídia Forense. Autenticidade de áudio. Codificador AMR. Dupla Compressão.

Análisis Forense de AMR Compressed-Domain para Detección de Doble Compresión Multimedia

Resumen

Para ser aceptada como prueba en el proceso criminal, una grabación de audio debe ser auténtica para preservar con la máxima fidelidad posible las conversaciones registradas y evitar errores de interpretación. El codificador AMR es un estándar mundial para compresión y transmisión de voz en las redes de telefonía móvil GSM 3G y 4G. Además, tal codificador es un estándar de almacenamiento de archivos de extensión AMR que utiliza el mismo algoritmo de compresión. Por su extenso uso en las redes móviles y alta disponibilidad en los modernos smartphones, el formato AMR ha sido encontrado en casos de exámenes de autenticidad de audio en que se cuestiona si hubo alguna manipulación. Tales exámenes forman parte del área de multimedia forense que está formada, entre otras técnicas, por la detección de doble compresión, o sea, determinar si un archivo AMR fue descomprimido y comprimido nuevamente. La detección de la doble compresión AMR es un problema complejo de ingeniería cuya solución aún está en curso. En líneas generales, si un archivo AMR tiene doble compresión, no es compatible con un original y puede haber sido modificado. Los trabajos publicados en la literatura sobre detección de doble compresión utilizan el archivo AMR decodificado en el dominio del tiempo como punto de partida para la extracción de características. En este artículo se propone un nuevo enfoque para el problema de la detección de doble compresión AMR en el que, en lugar de utilizar el audio decodificado, se utiliza su versión codificada para la extracción de características en el dominio de la compresión basadas en los coeficientes de predicción lineal (LP). Por medio de análisis estadístico de esas características, es posible mostrar que ellas pueden ser usadas para realizar la detección de doble compresión AMR de forma efectiva. De esta forma, las características en el campo de la compresión pueden considerarse un camino prometedor para la resolución del problema de la doble compresión AMR a través de redes neuronales artificiales.

Palabras Claves: Multimedia Forense. Autenticidad de audio. Codificador AMR. Doble Compresión.

References

3rd Generation Partnership Project (3GPP). AMR codec - Release 10, 2011. Available at: <http://www.3gpp.org/ftp/Specs/archive/26_series/26.104/>. Acessed: jan. 22 2018.

BATTIATO, S.; GIUDICE, O.; PARATORE, A. Multimedia Forensics: Discovering the History of Multimedia Contents. In: Proceedings of the 17th International Conference on Computer Systems and Technologies (CompSysTech), Palermo, Italy, p. 5-16, 2016.

ESQUEF, P. A. A.; APOLINÁRIO, Jr. A.; BISCAINHO, L. W. P. Edit Detection in Speech Recordings Via Instantaneous Electric Network Frequency Variations. IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, p. 2314-2326, dec. 2014.

FARID, H. Detecting Digital Forgeries Using Bispectral Analysis. AI Lab, MIT Tech. Rep. AIM-1657, Boston, USA, 1999. Available at: <ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-1657.pdf>. Acessed: jan. 22 2018.

GAROFOLO, J. S.; LAMEL, L. F.; FISHER, W. M.; FISCUS, J. G.; PALLETT, D.S.; DAHLGREN, L.; ZUE, V. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Philadelphia: Linguistic Data Consortium, 1993. Avaliable AT: http:// https://catalog.ldc.upenn.edu/LDC93S1. Acessed: jan. 22 2018.

GRIGORAS, C. Forensic Analysis of the Digital Audio Recordings - The Electric Network Frequency Criterion, Forensic Science International, vol. 136, Suppl. 1, p. 368-369, 2003.

GUPTA, S.; CHO, S.; KUO, C. C. Current Developments and Future Trends in Audio Authentication. IEEE Multimedia, vol. 19, no. 01, p. 50-59, jan. 2012.

HANILÇI, C.; KINNUNNEN, T. Source Cell-Phone Recognition from Recorded Speech using Non-Speech Segments. Digital Signal Processing, no. 35, p. 75-85, 2014.

HIÇSÖNMEZ, S.; SENCAR, H. T.; AVCIBAS, I. Audio Codec

Identification through Payload Sampling. In: Proc. Int. Workshop Inf. Forensics Secur., Iguacu Falls, Brazil, p. 1-6, 2011.

HIÇSÖNMEZ, S.; UZUN, E.; SENCAR, H. T.; Methods for Identifying Traces of Compression in Audio. In: Proc. 1st Int. Conf. Commun., Signal Process., Appl., Sharjah, United Arab Emirates, p. 1-6, 2013.

IKRAM, S.; MALIK, H. Digital Audio Forensics Using Background Noise. In: 2010 IEEE International Conference on Multimedia & Expo (ICME'2010), Singapore, p. 106-110, 2010.

JENNER, F.; KWASINSKI, A. Highly Accurate Non-Intrusive Speech Forensics for Codec Identifications from Observed Decoded Signals. In: Proc. 2012 IEEE International Conference on Acoustic, Speech and Signal Processing., Kyoto, Japan, p. 1737-1740, 2012.

KAJSTURA, M.; TRAWINSKA, A.; HEBENSTREIT, J. Application of the Electrical Network Frequency (ENF) Criterion: a Case of a Digital Recording. Forensic Science International, nº 155, p. 165-171, feb. 2005.

KORYCKI, R. Detection of Montage in Lossy Compressed Digital Audio Recordings. Archives of Acoustics, vol. 39, no. 1, p. 67-52, feb. 2014.

LIU, Q.; SUNG, A. H.; QIAO, M. Detection of Double MP3 Compression. Cognitive Computation., vol.2, no. 4, p. 291-296, 2010.

LUO, D.; LUO, W.; YANG, R.; HUANG, J. Compression History Identification for Digital Audio Signal. In: Proc. 2012 IEEE International Conference on Acoustic, Speech and Signal Processing., Kyoto, Japan, p. 1732-1736, 2012.

LUO, D.; LUO, W.; YANG, R.; HUANG, J. Identifying Compression History of Wave Audio and its Applications. ACM Trans. Multimedia Comput., Commun., Appl., vol. 10, no. 3, p. 30-1 - 30-19, 2014.

LUO, D.; YANG, R.; HUANG, J. Detecting Double Compressed

AMR Audio using Deep Learning. In: 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP'2014), Florence, Italy, p. 2688-2692, 2014.

LUO, D.; YANG, R.; LIN, B.; HUANG, J. Detection of Double Compressed AMR Audio Using Stacked Autoencoder. IEEE Transactions on Information Forensics and Security, vol. 12, no.2, p. 432-444, 2017.

MALIK, H. Acoustic Environment Identification and Its Applications to Audio Forensics. IEEE Transactions on Information Forensics and Security, vol. 8, no. 11, p. 1827-1837, nov. 2013.

PETRACCA, H.; SERVETTI, A.; DE MARTIN, J. C. Low-Complexity Automatic Speaker Recognition in the Compressed GSM AMR Domain In: Proc. IEEE Int. Conf. Multimedia and Expo (ICME), Amsterdam, Netherlands, 2005.

PFEIFFER, S.; VINCENT, T. Survey of Compressed Domain Audio Features and their Expressiveness. In: Proc. SPIE Conf. Electronic Imaging, San Francisco, USA, p. 133-147, 2003.

QIAO, M.; SUNG, A. H.; LIU, Q. Revealing Real Quality of Double Compressed MP3 Audio. In: Proc. Int. Conf. Multimedia, Florence, Italy, p. 1011-1014, 2010.

REIS, P. M. G. I.; COSTA, J. P. C. L.; MIRANDA, R. K.; DEL GALDO, G. ESPRIT-Hilbert-based Audio Tampering Detection with SVM Classifier for Forensic Analysis Via Electrical Network Frequency. IEEE Transactions on Information Forensics and Security, vol. 12, no. 4, p. 853-864, apr. 2017.

RODRIGUEZ, D. P. N.; APOLINÁRIO, J. A.; BISCAINHO, L. W. P. Audio Authenticity: Detecting ENF Discontinuity with High Precision Phase Analysis. IEEE Transactions on Information Forensics and Security, vol. 5, no. 3, p. 534-543, sep. 2010.

ROMERO, D.; WILSON, C. Y. Automatic Acquisition Device

Identification from Speech Recordings. In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'2010), 2010, Texas, USA, p. 1806-1809, 2010.

SHEN, Y.; JIA, J.; CAI, L. Detecting Double Compressed AMR-format Audio Recordings. In: Proc. 10th Phonetics Conf. China (PCC), China, p. 1-5, 2012.

SU, H.; GARG, R.; HAJJ-AHMAD, A.; WU, M. ENF Analysis on Recaptured Audio Recordings. In: 2013 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Canada, p. 3018-3022, 2013.

YANG, R.; QU, Z.; HUANG, J. Exposing MP3 Audio Forgeries Using Frame Offsets. ACM Transactions on Multimedia Computing, Communications and Applications, vol. 8, no. S2, p. 35:1-35:20, sep. 2012.

YANG, R.; SHI, Y. Q.; HUANG, J. Defeating Fake-quality MP3. In: Proc. ACM Workshop Multimedia Security, New Jersey, USA, p. 117-124, 2009.

YANG, R.; SHI, Y. T.; HUANG, J. Detecting Double Compression of Audio Signal. In: Proc. SPIE Conference on Media Forensics and Security II, San Jose, USA, vol. 7541, p. 1-10, 2010.

ZAKARIAH, M.; KHAN, M. K.; MALIK, H. Digital Multimedia Audio Forensics: Past, Present and Future. Multimedia Tools and Applications, vol. 77, p. 1009-1040, 2017.