Methods for Identifying Traces of Compression in Audio

Samet Hicsonmez
Computer Engineering Department,
TOBB University of
Economics and Technology,
Ankara, Turkey
Email: shicsonmez@etu.edu.tr

Erkam Uzun
Computer Engineering Department,
TOBB University of
Economics and Technology,
Ankara, Turkey
Email: euzun@etu.edu.tr

Husrev Taha Sencar
Computer Engineering Department,
TOBB University of
Economics and Technology,
Ankara, Turkey
Email: htsencar@etu.edu.tr

***Abstract*—Compression history of an audio may reveal very useful information when traces of tampering has to be investigated or quality of an audio has to be evaluated. Motivated by this, we introduce two methods that can discriminate between single and double compressed audio and can identify compression codec and bit rate of an audio. The first method utilizes audio quality measures to realize this and operates on decoded audio. The second method, alternatively, works on coded audio, effectively the audio bit stream, and characterizes randomness and chaotic properties of the bit stream to achieve these tasks. Unlike the existing work in the literature, which are proposed mainly for MP3 encoded audio, both methods can be applied to all encoding formats. Extensive tests have been performed to test the performance of both methods under various settings. Results show that both methods can be very reliably used to obtain information on compression history of an audio.**

I. Introduction

With powerful editing tools, tampering of an audio without leaving any audible traces or artifacts has become an easy task. In practice, audio tampering may be as blatant as deletion or insertion of audio fragments or splicing segments of audio together and may be as subtle as resampling of the audio. Since there are many such ways to tamper an audio, the ability to investigate audio recordings for evidence of editing and tampering is very crucial. Over time, many techniques have been developed to achieve this goal by identifying traces of various signal processing operations, analyzing harmonic consistency and examining background noise characteristics.

Today, audio is almost always stored and transmitted in some coded form; therefore, any type of audio tampering will require encoded audio to be first decoded, processed and later encoded again with one of many codecs. The most important goal of encoding is compression, which mainly determines the data rate of coded audio. Since codecs are designed to utilize different techniques to code raw audio samples and each encoder operates at different compression levels, consecutive encoding of an audio with different codecs or a change in the data rate of an audio essentially translates to multiple compression of the audio. Therefore, being able to determine compression history of an audio or, at least, identify doubly compressed audio plays a vital role in determining whether an audio is in the format that it was initially recorded at or it has undergone some processing. Such a capability also makes it possible to evaluate the quality of an audio and detecting fake-quality ones, i.e., low bit rate audio files transcoded at higher bit rates pretending to be in high quality [1].

There is only a limited number of works that aim at determining compression history of an audio and identifying doubly compressed audio [2] [3] [4] [5] [6]. Yang et al. [3] introduced a method to identify fake-quality MP3 files that are generated by up-transcoding lower quality MP3 files. They observed that the number of MDCT coefficients with small values in true quality MP3 files is higher than it is in fake-quality ones. Based on this observation, they proposed to use the number of MDCT coefficients with ± 1 values in identifying up-transcoded MP3 files. In [6], an improved version of the method in [3] is introduced to determine not only up-transcoded MP3 audio but also down-transcoded MP3 audio. The method utilizes both zero and non-zero MDCT coefficients to obtain several statistics from which 117 features are extracted for classification.

In [2], authors proposed a method to determine whether a given uncoded audio was previously coded with WMA or MP3 codecs and to estimate coding bit-rate, if the audio was coded. The method is based on detecting traces of compression through analysis of MDCT coefficients. This is realized by comparing the average number of zero-valued MDCT coefficients and the mean values of MDCT coefficients grouped into 24 bins. In [7], a method for detecting MP3 coded audio forgeries, like insertion, deletion and substitution, is proposed by examining frame offsets, which required identification of quantization characteristics. For this they utilized two important features, namely, the number of non-zero spectral coefficients and the properties of sorted spectral coefficient values.

All these works, report highly accurate performance results. However, they all have limited applicability when considered in the context of detecting audio tampering as they are primarily focus on MP3 encoding. Further, since at their core these techniques utilize specifics of MP3 encoding they cannot be trivially extended to other encoding techniques. In practice, tampering concerns audio samples captured by various sources like voice recorders or samples recorded on landline, cell phone and VoIP communication networks. Most typically, tampered audio is encoded with one of the popular encoding formats such as WMA, AAC and FLAC in addition to MP3.

In this paper, we propose two methods to obtain information on compression history of a given audio. Both methods take into account two different scenarios. The first scenario concerns with singly compressed audio, by which we refer to an audio encoded with a compression codec. The second one considers doubly compressed coded audio, which refers to audio encoded twice by the same or different codecs but most generally with a change in its data rate. The goal in this scenario is to determine the codec used prior to transcoding, which provides information on the initial compression level of the audio. The two methods differ depending upon whether the analysis has to be performed on coded or uncoded audio. The first method works on decompressed (WAV formatted) audio to determine the codec used for encoding. The second method, on the other hand, assumes that only the coded audio is available for analysis.

In the following section, both of the methods are explained in detail. Section III provides results for the carried out experiments and presents performance results for the two methods. Lastly, our remarks are given in Section IV.

978-1-4673-2821-0/13$31.00 © 2013 IEEE

II. METHODOLOGY

The methods described in this section can be used for tracing the compression history of both single and double encoded audios. In this context, the first method utilizes audio quality measures extracted from encoded (i.e., compressed) and decoded (i.e., uncompressed) audio samples to differentiate between different levels of compression. The second method, unlike the first one, attempts to make a decision only through analysis of coded audio, effectively the bit-stream, without decoding it or using any encoding metadata. It is essentially based on characterization of encoding by measuring the inherent chaotic and randomness characteristics of encoded data. In [8], we initially proposed this method for fast and reliable identification of codec used in encoding of an audio, and here we investigate its capability to distinguish characteristics of compression. More specifically, we test its ability to identify doubly compressed audio.

A. Method I

To measure and assess the impact of compression on audio samples, we deploy audio quality measures. These measures are primarily used to diagnose problems in the quality of an audio by detecting and quantifying the distortion due to encoding. Furthermore, it has been demonstrated that such measures can be successfully employed in audio steganalysis, where an audio is modified in much more subtle ways than compression and the goal is to determine whether or not an audio has undergone information embedding operation [9]. Essentially, quality of an audio that is subjected to some form of processing needs to be evaluated with respect to a reference signal. In conventional audio quality measurement, this reference is taken as the original, unmodified audio. However, when it is not feasible to access the original audio, an alternative approach is to generate a reference that approximates the original. Most commonly this is done by denoising the processed audio with the assumption that both the original and its processed version will yield similar signals.

When an audio needs to be examined for traces of tampering, in fact, presence of an original is in question. Therefore, to be able link audio quality measurements to compression history of an audio, a reference needs to be obtained blindly. There are several methods to denoise a signal, which may be based on wavelet shrinkage, independent component analysis, maximum likelihood and discrete wavelet transform. In our method, we use a wavelet based de-noising approach [10].

In total, 21 quality measures, relevant to perceptual, time and frequency domains are extracted. These measures are then combined into a feature vector, which is used for classification of audios with different compression history. Each audio quality measure is computed on a frame of audio that lasts 20-100ms, and the final value is obtained as the average of measurements from each frame. The formulation for each quality measure and the corresponding frame size is given in Table I. In the table, *y(i)* and *x(i)* represents the original and the reference signal, respectively. $B_x(i)$, $B_y(i)$, $S_x(i)$, $S_y(i)$ and *M(i)* represents Bark spectra in the i'th critical band and perceptible distortion, respectively. And, a, c, Q and $X(w), Y(w)$, respectively, represent LPC coefficient vector, cepstral coefficients, phase spectrum and magnitude spectrum. In what follows next, we briefly describe the quality measures used in our tests.

TABLE I
AUDIO QUALITY MEASURES

Measure	Frame Size	Definition
		Perceptual Domain Measures
BSD	60 ms	$\sum_{i=1}^{C}[B_x(i)-B_y(i)]^2$ $C: CriticalBandSize$
MBSD	80 ms	$\sum_{k=1}^{C} M(i)\lvert S_x(i)-S_y(i)\rvert$
EMBSD	20 ms	$\sum_{i=1}^{15} Max\{\lvert S_x(i)-S_y(i)\rvert-Th(i),0\} * \lvert S_x(i)-S_y(i)\rvert$
PAQM	32 ms	Gives a distortion value between 0 and 6,5 for audio signals.
PSQM	32 ms	Gives a distortion value between 0 and 6,5 for speech signals.
WSSD	40 ms	$\sum_{k=1}^{3} 6w(k)\left\{\begin{array}{l}[X(k+1)-X(k)]-\\ [Y(k+1)-Y(k)]\end{array}\right\}^2$
MNB	60 ms	Calculated with different time-frequency structures
		Time Domain Measures
SNR	20 ms	$10\log_{10}\frac{\sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N}(x_i-y_i)^2}$
SNRseg	20 ms	$\frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{i=K_m}^{K_m+K-1} x_i^2}{\sum_{i=K_m}^{K_m+K-1}(x_i-y_i)^2}$
CZD	40 ms	$\frac{1}{K}\sum_{i=0}^{K-1}\left(1-\frac{2min(x_i,y_i)}{x_i-y_i}\right)$
		Frequency Domain Measures
LLR	60 ms	$\log\left(\frac{a_x^T R_y a_x}{a_y^T R_y a_y}\right)$
LAR	60 ms	Depends on partial correlation coefficients
ISD	100 ms	$\int_{-\pi}^{\pi}\left(\log\frac{Y(w)}{X(w)}+\frac{X(w)}{Y(w)}-1\right)\frac{dw}{2\pi}$
COSH	100 ms	$\int_{-\pi}^{\pi}\left(\frac{1}{2}\left(\frac{Y(w)}{X(w)}+\frac{X(w)}{Y(w)}\right)-1\right)\frac{dw}{2\pi}$
CD	20 ms	$\left[\begin{array}{l}[c_x(0)-c_y(0)]^2+\\ 2\sum_{k=1}^{K}[c_x(k)-c_y(k)]^2\end{array}\right]^{1/2}$
SP	40 ms	$\frac{1}{K}\sum_{w=1}^{K}\lvert Q_x(w)-Q_y(w)\rvert^2$
SPM	20 ms	$\frac{1}{K}\left(\begin{array}{l}\lambda\sum_{w=1}^{K}\lvert Q_x(w)-Q_y(w)\rvert^2+\\ (1-\lambda)\sum_{w=1}^{K}\lvert\lvert X(w)\rvert-\lvert Y(w)\rvert\rvert^2\end{array}\right)$ $\lambda=0.025$
STFT	60 ms	$R(\rho,\theta)=\int_x\int_y S(\tau,\omega)\delta\left(\begin{array}{l}\tau cos\theta+\\ \omega sin\theta-\\ \rho\end{array}\right)d\tau d\omega$

***Perceptual Domain Features**:*

Perceptual Audio Quality Measure (PAQM): A model of human auditory system is simulated in calculation of this measure. A quantity between 0 and 6,5 is determined as the quality measure.

Perceptual Speech Quality Measure (PSQM): Quantifies voice quality of speech codecs operating in the frequency band 300Hz - 3400 Hz. Similar to PAQM, it takes values in 0-6,5 range [11].

Bark Spectral Distortion (BSD): Estimates the overall distortion by using the average Euclidean distance between loudness vectors of the given audio and the reference audio [12].

Modified Bark Spectral Distortion (MBSD): It is computed similar to BSD with the only difference that it also takes into account a noise masking threshold to estimate the audible and inaudible distortions [13].

Enhanced Modified Bark Spectral Distortion (EMBSD): Only the first 15 loudness components are involved in computation of the MBSD [13].

Measuring Normalizing Blocks (MNB): Used for estimating speech quality by transforming speech signal into approximate loudness domain through frequency warping and logarithmic scaling [14].

Weighted Slope Spectral Distance (WSSD): Reflects phonetic differences between spectral slopes in speech recognition applications [15].

Time Domain Features*:*

Signal-to-Noise Ratio (SNR): A measure of signal strength relative to the background noise.

Segmental Signal-to-Noise Ratio (SNRseg): Average value of the SNR calculated in each audio frame.

Czekanowski Distance (CZD): Correlation based measure that compares two signals in time domain.

Frequency Domain Features*:*

Log-Likelihood Ratio (LLR): The LLR, also known as Itakura Distance, considers an all-pole linear predictive coding (LPC) model on speech segments [16].

Log-Area Ratio (LAR): Another LPC-based distance that uses partial correlation coefficients.

Itakura-Saito Distance (ISD): Represents the discrepanciency between the power spectrums of the given audio and the reference signal.

COSH Distance (COSHD): Symmetric version of Itakura-Saito distance.

Cepstral Distance Measure (CD): The difference between cepstral coefficients of the given audio and the reference signal.

Spectral Phase Distortion (SP) and Spectral Phase Magnitude Distortion (SPM): Represents the phase distortions between the two signals.

Short Time Fourier-Radon Transformation (STFRT): Defined as the mean-square distance between Radon transforms of the short time Fourier transform (STFT) of two signals.

B. Method II

In [8], we introduced a new technique for characterization of encoded data and demonstrated that it can distinguish between many codecs used in encoding of speech and music. The underlying idea of the method is that factors influencing the design of a codec, like encoding technique, compression level, quality, and complexity of a codec, has a combined effect that reveals itself in the chaotic and randomness properties of the coded bit-stream. To capture such characteristics, a few kilobytes of data is randomly carved out from the encoded audio and a number of statistics have been proposed. These statistics are then used as features to build a multi-class classification system.

Features used for characterizing coded audio are grouped into two main categories. First group comprises randomness features that are determined by statistical analysis of sampled byte streams in both time and frequency domain. In time domain, simple statistics like mean, variance, entropy, auto-correlation function (first 21 coefficients) and higher order statistics such as bicoherence, skewness and kurtosis are calculated. In frequency domain, the spectrum is divided into four equal subbands and mean, variance and skewness in each subband are computed as features Overall, there are 39 randomness features with 27 computed in time domain and 12 in frequency domain.

The other group includes chaotic features such as *Lyapunov exponents* (LE) and *false neighbors fraction* (FNF). Concerning FNF, the fraction of false nearest neighbors, average size of the neighborhood and average of the squared size of the neighborhood for first 5th embedding dimensions are calculated. As LE type features, logarithm of the stretching factor for first 11 iterations are calculated. With the addition of these 26 chaotic type features, we end up having a 65 dimensional feature vector.

Tests performed considering 16 very popular codecs used in encoding of speech and music at data rates ranging from 8 kbps to 128 kbps show that an identification accuracy higher than 96% can be achieved. Since these results establish the fact that singly-encoded audio, at difference compression levels, can be reliably discriminated, in this work, we take it one step further and test the technique's ability to identify doubly-encoded audio.

III. Experiments

Experiments are conducted on four data sets. First set consists of 500 speech samples taken from VoxForge speech corpus [17]. Each sample in this set is 1-13 seconds long and has a data rate of 256 Kbps. Second data set consists of 1000 five second long music samples taken from 500 different songs across various genres, recorded at CD quality (1411 Kbps). The two samples taken from each song are obtained by pulling out two randomly determined non-overlapping and non-contiguous segments. In the rest of the paper, the first data set will be referred to as Speech data set, and the second as Music data set.

In some of the tests, we used two more data sets to assess the generality of the results. One of them consists of 1000 music samples provided as part of Marsyas project [18]. These samples are all 22KHz, mono, 30 seconds long and have 353 Kbps bit rate. Last data set consists of 4000 music samples taken from 4000 different songs, and it is generated in a manner similar to Music data set. These latter two data sets will be hereafter referred to as Marsyas and Music-II data set.

We performed two groups of tests considering both methods. First group of tests are conducted on uncoded audio to determine the performance of Method I in identifying compression bit rate of single and double compressed audio. Second group of tests are designed to measure the ability of Method II in identifying compression bit rate of double-compressed audio, and tests are performed on coded audio. In all tests, we use a standard machine learning technique, support vector machine (svm) implemented in the Libsvm package [19] with radial basis kernel for classification. In all cases, half of the samples are used for training the other half is used for testing.

A. Tests on Uncoded Audios

Tests performed on single and double compressed audio are discussed separately in the following subsections.

1) Single Compression Tests:

Bit-Rate Test*:* We used Music data set in experiments to determine whether a given WAV formatted audio is an original (uncompressed) audio or an audio compressed at a fixed rate. Audio samples in this data set have been compressed and decompressed by a mix of codecs such as MP3, AAC, WMA and OGG, at 32 Kbps, 40 Kbps, 48 Kbps, 56 Kbps and 64 Kbps compression rates. Results associated with the experiments are given in Table II. Average identification accuracy in differentiating uncompressed audio from compression audio at different bit rates is calculated as 98.62%.

TABLE II
CLASSIFICATION ACCURACIES FOR DISCRIMINATING BETWEEN ORIGINAL AND SINGLE COMPRESSED WAV FORMATTED AUDIO

Bit Rate	32 kbps	40 kbps	48 kbps	56 kbps	64 kbps
Accuracy	100%	99.9%	100%	99.2%	94.0%

To determine the method's ability to identify the compression bit rate, a second experiment is performed. In this case, rather than deploying binary classification, a multi-class classifier is built. Table III presents the corresponding confusion matrix for this experiment. The average accuracy of this experiment is 94.77%.

TABLE III
CONFUSION MATRIX FOR IDENTIFYING COMPRESSION BIT-RATES OF SINGLE COMPRESSED AUDIO.

Bit-Rate	Wav	32 kbps	40 kbps	48 kbps	56 kbps	64 kbps
Wav	89.6%	*	*	*	*	10.2%
32 kbps	*	98.0%	*	*	*	*
40 kbps	*	*	98.0%	*	*	*
48 kbps	*	*	*	99.0%	*	*
56 kbps	*	*	*	*	96.4%	*
64 kbps	12.4%	*	*	*	*	87.6%

Encoder Test: In these experiments, we used Speech data set to detect whether a given WAV file has been compressed with AAC, AMR, G.729, GSM 6.10, GSMWAV, MP3 codecs or not compressed at all. (It should be noted that AMR, G.729, GSM 6.10, GSMWAV codecs are designed for speech coding.) In the experiments, compression bit rate for each codec is set to its default bit rate. Compression bit rates are different for all codecs and they vary between 10kbps and 128 kbps. The results associated with these experiments, where a binary classifier is built for each encoder, is given in Table IV. Average accuracy in discriminating audio compressed with these codecs from the uncompressed, original, audio is found to be 98.80%.

TABLE IV
CLASSIFICATION ACCURACIES FOR DISCRIMINATING WAV FORMATTED UNCOMPRESSED AUDIO FROM AUDIO ENCODED WITH VARIOUS CODECS AT DIFFERENT BIT RATES

Codec	Aac	Amr	G729	Gsm 6.10	GsmWav	Mp3
Accuracy	94.0%	100%	98.8%	100%	100%	100%

We also extended this experiment to a multi-class classification scenario to test method's ability to distinguish audio encoded with different codecs. The confusion matrix associated with this experiment are given in Table V. Classification accuracy corresponding to this experiment is found to be 96.60%.

TABLE V
CONFUSION MATRIX FOR IDENTIFYING ENCODER OF SINGLE COMPRESSED AUDIO.

Codec	Aac	Amr	G.729	Gsm 6.10	GsmWav	Mp3
Aac	98.0%	*	*	*	*	*
Amr	*	91.2%	*	8.8%	*	*
G.729	*	*	100%	*	*	*
Gsm 6.10	*	8.8%	*	91.2%	*	*
GsmWav	*	*	*	*	100%	*
Mp3	*	*	*	*	*	99.2%

2) Double Compression Tests: In the previous tests, audio samples are compressed using AAC, AMR, G.729, GSM 6.10, GSMWAV and MP3 codecs. In these tests, these audio samples are re-compressed using AAC and MP3 codecs at relatively high bit rates. The purpose of the first test is to identify whether a given AAC or MP3 compressed WAV formatted audio is in fact single or double decompressed audio. Results corresponding to this experiment are given in Table VI. Regardless of whether the second codec is AAC or MP3, it can be seen that an average accuracy above 95% can be obtained in classifying single and double coded audio.

TABLE VI
CLASSIFICATION ACCURACIES FOR DISCRIMINATION OF SINGLE AND DOUBLE COMPRESSED AUDIO

Second Codec	First Codec					
	Aac	Amr	G.729	Gsm 6.10	GsmWav	Mp3
Aac	*	100%	96.8%	100%	99.2%	98.2%
Mp3	90.8%	99.6%	90.4%	100%	98.0%	*

Another experiment is performed to test the ability to identify the first compression codec of a double compressed audio. For this purpose, multi-class classifiers are built to differentiate between audio compressed with one of the above codecs followed by re-compression. Tables VII and VIII provide the confusion matrices associated with these experiments, where MP3 and AAC codecs are, respectively, used for re-compression. The results show that the proposed method has an accuracy of around 90% in identifying the codec used during initial compression step.

TABLE VII
CONFUSION MATRIX FOR DETECTING INITIAL COMPRESSION CODEC OF DOUBLE COMPRESSED AUDIOS WITH AAC CODEC. (OVERALL ACCURACY 91.13%)

Codec	Aac	Amr	G.729	Gsm 6.10	GsmWav	Mp3
Aac	92.0%	*	5.2%	*	*	*
Amr	*	91.6%	*	8.4%	*	*
G.729	*	*	97.6%	*	*	*
Gsm 6.10	*	28.4%	*	71.2%	*	*
GsmWav	*	*	*	*	98.8%	*
Mp3	4.4%	*	*	*	*	95.6%

TABLE VIII
CONFUSION MATRIX FOR DETECTING INITIAL COMPRESSION CODEC OF DOUBLE COMPRESSED AUDIOS WITH MP3 CODEC. (OVERALL ACCURACY 88.80%)

Codec	Aac	Amr	G.729	Gsm 6.10	GsmWav	Mp3
Aac	97.2%	*	*	*	*	*
Amr	*	71.6%	*	25.2%	2.8%	*
G.729	*	*	97.6%	*	*	*
Gsm 6.10	*	23.6%	*	74.4%	*	*
GsmWav	*	*	*	*	95.2%	*
Mp3	2.8%	*	*	*	*	96.8%

B. Tests on Coded Audio

Two group of tests are performed on coded audio. The first group concerns double compressed audio. In these tests, we aim to determine the performance of Method II in discriminating between single and double compressed audio and identifying the initial compression rate (prior to transcoding). (Since our earlier work, [8], covers coding history identification for single coded audios using this method, those tests are not repeated here.) The second set of experiments focus on detection of MP3 files with fake quality, i.e., low quality audio re-encode at higher bit-rates to give the impression that they are of high quality. Here the emphasis on MP3 is due to the fact that MP3 has been the most commonly used codec in distributing and sharing music.

1) Double Compression Tests: Experiments are performed considering the fact that in a tampering scenario, audio will most likely be re-encoded with one of the codecs like MP3, AAC and WMA due to extensive support for those codecs on various platforms. We first compress raw audio samples in the Speech data set with PSTN codecs like a-law and u-law, GSM codecs like AMR and GSM 6.10, and VoIP codecs like G.726, G.729 and Speex. It must be noted that all these codecs provide compression bit rates ranging from 64 Kbps to 8 Kbps. Then all these coded audios are re-encoded with MP3 at 64 Kbps and with AAC at 128 Kbps.

To test the method's ability to distinguish double compressed audios, which are first coded with one of the mentioned PSTN, GSM and VoIP codecs and later re-encoded with MP3 and AAC codecs, from single compressed audios encoded with MP3 or AAC codecs, we combined all single and double compressed audios in separate classes and built a binary classification system. Results show that the method can discriminate between single and double coded audios with an accuracy of 95.8% and 82.8% when AAC and MP3 are used as the transcodecs, respectively. Examining the results, where MP3 codec is used for re-compression, we observed that confusions are mainly due to VoIP codecs. Repeating the same experiment with the exclusion of the VoIP codecs, we observed that the accuracy increased to of 99.5%. Considering the MP3 transcoding case, we performed several binary classification experiments to demonstrate the discrimination between single and double compressed audios, where seven different codecs are used for initial encoding. Classification results corresponding to the these experiments are given in Table IX.

TABLE IX
BINARY CLASSIFICATION RESULTS FOR SINGLE AND DOUBLE COMPRESSED AUDIO WHERE MP3 CODED IS USED FOR DECOMPRESSION

Transcodec	First Codec						
	A-law	U-law	Amr	Gsm	G.726	G.729	Speex
MP3	99.8%	99.4%	98.8%	99.6%	63.8%	68.4	54.4%

2) Fake Quality MP3 Tests: MP3 file format still remains to be the most popular audio file format used for encoding music content, and fake quality MP3 files are generated by re-encoding lower bit rate files at a higher rate. It also needs to be emphasized that existing work in the literature that aim at identification of compression history of an audio has mainly focused on this scenario. To detect fake quality MP3 files, we conducted extensive experiments using the three music data sets, namely, Music, Marsyas and Music-II data sets. We choose 32, 40, 48 and 56 Kbps as bit rates of original lower bit rate MP3 files and highly popular bit rates 192, 256 and 320 Kbps as up-transcoding bit rates. For testing the performance of the method, we first encode all uncoded samples of all data sets using 32, 40, 48, 56, 192, 256 and 320 Kbps bit rate options. Then, 32, 40, 48 and 56 Kbps coded MP3 files are decoded and re-encoded at 192, 256 and 320 Kbps bit rates. This yielded two types of encoded MP3 files where one group consists of samples double compressed at a fake quality and the other including single compressed high quality samples.

Using these two datasets we constructed a binary classification system to identify whether or not a given MP3 file has fake quality. Corresponding test results are presented in Tables X, XI and XII for all the data sets where the final compression bit rate is, respectively, set at 192, 256 and 320 Kbps. In the tables, first column shows the initial and second compression bit rates for transcoded MP3 files. Since all these tests establish binary classifications, second bit rate also indicates the bit rate for true quality MP3 files. Average accuracies for identification of fake quality MP3 files in each data sets are determined to be as 98.7%, 98.4% and 99.9%.

To ensure that classification accuracies are not biased by the initial compression bit rate, all the fake quality MP3 files at different initial compression rates are combined into a single class and tested against true quality MP3 files. (In this case, when training the classifier, the number of true quality and fake quality MP3 files are still kept equal by selecting equal sized subsets of samples from each fake quality MP3 group.) The last column in each table gives the results corresponding to this test case. It can be seen that even under this setting, average identification accuracies are above 96.5% for each data set.

TABLE X
RESULTS FOR CLASSIFICATION OF FAKE AND TRUE QUALITY MP3 FILES ENCODED AT 192 KBPS.

	Accuracy (%)		
First-Second Bitrate	Music-I (44Khz)	Marsyas (22Khz)	Music-II (44Khz)
32-192	98.7	98.4	99.9
40-192	98.6	98.3	99.87
48-192	98.7	98.3	99.77
56-192	97.8	97.8	99
All-192	98.1	98.3	99.4

TABLE XI
RESULTS FOR CLASSIFICATION OF FAKE AND TRUE QUALITY MP3 FILES ENCODED AT 256 KBPS.

	Accuracy (%)		
First-Second Bitrate	Music-I (44Khz)	Marsyas (22Khz)	Music-II (44Khz)
32-256	98.6	98.4	99.6
40-256	97.5	98.3	99.97
48-256	97.9	97.9	99.97
56-256	99.2	99.2	98.72
All-256	98.3	97.8	98

TABLE XII
RESULTS FOR CLASSIFICATION OF FAKE AND TRUE QUALITY MP3 FILES ENCODED AT 320 KBPS.

	Accuracy (%)		
First-Second Bitrate	Music-I (44Khz)	Marsyas (22Khz)	Music-II (44Khz)
32-320	96.8	96.7	99.42
40-320	96.5	95.8	99.95
48-320	97.9	97.1	99.95
56-320	99	98.7	98.05
All-320	97.4	96.5	97.9

IV. CONCLUSION AND FUTURE WORK

Two methods are proposed to identify traces of compression in encoded audio. Proposed methods can be used to identify the codec used in compression of an audio and its compression bit rate, to discriminate between single and double compressed audio, and to determine the initial compression codec, along with its compression bit rate, of a double compressed audio.

Results show that Method I is able to classify a given WAV file as single or double compressed with an accuracy around 95%. Accuracies for codec identification from single and double compressed audio are measured to be around 97% and 90%, respectively. Moreover, Method I has the ability to detect the bit rates of single compressed audio with various codecs such as MP3, AAC, WMA and OGG with an accuracy higher than 95%. Results also show that Method II is able to determine whether a given audio is double or single compressed with an accuracy above 95% as long as transcoding is performed at relatively high bit rates. In a similar manner, the method can identify fake quality MP3 files with an accuracy close 98%. Overall, these results are among the highest reported in the literature

ACKNOWLEDGMENT

This material is based upon work supported by the Scientific and Technological Research Council of Turkey under research grant number 110E049.

REFERENCES

[1] "Ifpi digital music report," 2009. [Online]. Available: http://www.ifpi.org/content/library/DMR2009.pdf

[2] R. Y. Da Luo, Weiqi Luo and J. Huang, "Compression history identification for digital audio signal," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ISASSP)*, 2012, pp. 1733–1736.

[3] R. Yang, Y.-Q. Shi, and J. Huang, "Defeating fake-quality mp3," in *Proceedings of the 11th ACM workshop on Multimedia and security*, 2009, pp. 117–124.

[4] R. Yang, Y. Q. Shi, and J. Huang, "Detecting double compression of audio signal," in *SPIE Conference on Media Forensics and Security*, 2010.

[5] Q. Liu, A. Sung, and M. Qiao, "Detection of double mp3 compression," *Cognitive Computation*, vol. 2, pp. 291–296, 2010.

[6] M. Qiao, A. H. Sung, and Q. Liu, "Revealing real quality of double compressed mp3 audio," in *Proceedings of the international conference on Multimedia*, ser. MM '10, 2010, pp. 1011–1014.

[7] R. Yang, Z. Qu, and J. Huang, "Detecting digital audio forgeries by checking frame offsets," in *Proceedings of the 10th ACM workshop on Multimedia and security*, 2008, pp. 21–26.

[8] S. Hicsonmez, H. Sencar, and I. Avcibas, "Audio codec identification through payload sampling," in *Information Forensics and Security (WIFS), 2011 IEEE International Workshop on*, 2011, pp. 1–6.

[9] H. Özer, B. Sankur, N. D. Memon, and I. Avcibas, "Detection of audio covert channels using statistical footprints of hidden messages," *Digital Signal Processing*, vol. 16, no. 4, pp. 389–401, 2006.

[10] D. L. Donoho and I. M. Johnstone, "Ideal denoising in an orthonormal basis chosen from a library of bases," *Comptes Rendus Acad. Sci., Ser. I*, vol. 319, pp. 1317–1322, 1994.

[11] "Perceptual speech quality measure." [Online]. Available: http://www.vocal.com/speech-coders/perceptual-speech-quality-measure-psqm

[12] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *Selected Areas in Communications, IEEE Journal on*, vol. 10, no. 5, pp. 819 –829, 1992.

[13] W. Yang, "Enhanced modified bark spectral distortion (embsd): An objective speech quality measure based on audible distortion and cognition model," Ph.D. dissertation, 1999.

[14] S. Voran, "Objective estimation of perceived speech quality-part i: Development of the measuring normalizing block technique," *IEEE Trans. on Speech and Audio Process*, pp. 371 – 382, 1999.

[15] B. Hanson and H. Wakita, "Spectral slope distance measures with linear prediction analysis for word recognition in noise," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 35, no. 7, pp. 968 – 973, 1987.

[16] F. Itakura, "Readings in speech recognition," A. Waibel and K.-F. Lee, Eds., 1990, ch. Minimum prediction residual principle applied to speech recognition, pp. 154–158.

[17] "Voxforge speech corpus." [Online]. Available: http://voxforge.org/

[18] G. Tzanetakis, G. Essl, and P. Cook, "Audio analysis using the discrete wavelet transform," *Proc. Conf. in Acoustics and Music Theory Applications*, 2001.

[19] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.