THINK INDIA JOURNAL

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

Audio Post-Processing Identification Using MFCC and LPC Feature

Loveneet Kaur

Asst. Professor, School of Computer Science and Engineering, Lovely

Professional University, Phagwara, Punjab, India
mangat.loveneet91@gmail.com

ShravankumarSauda

Research Scholar, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India.
shravan.ks9@gmail.com

Abstract

The vital communication medium, audios area unit simply changed or tampered throughout transmission; so the credibleness of audios is of high importance. This paper in the main introduces the technique for noticing the audio post-processing supported features of audio; the Support Vector Machine (SVM) is used for categorization throughout the detection. within the planned technique, Mel Frequency CepstralConstant (MFCC)[1] and also the Linear Prediction Coding (LPC)[2]of host audio area unit is measured as audio choices to which SVM is applied in order to protect the audios reputation. Results have shown that the primarily based technique of the planned audio feature cannot verify the credibility of the audio speed alone.Nevertheless, even the sleuthing of various kinds of post-processing operations has a major impact[3].

***Keywords:*Audio Feature;Audio Post-processing Detection; Support Vector Machine (SVM);Linear Prediction Coding (LPC);Mel Frequency Cepstral Coefficient (MFCC).**

1. Introduction

The dramatic growth in access to multiple new formed communication inventions makes it clear that multimedia has animportant role to play in criminal prosecution [4]. Audio post-processing methodologies, however, have emerged in an endless stream with the development of technologies.Different types of existing audio editing or polishingsoftware(s) can be accessed to perform different post-processing operations on audio clips, particularly after the removal and replacement of audio clips. Hence the quality of audios is highly vital.

A lot of research on encryption of audio signals has been done in recent years and several methods have been suggested to identify audio forgeries and the subsequent post-processing. Malik and Farid[5] found that contrasts in the estimated reverberation could be utilized in the forensic and ballistic, so the author(s) proposed a method for modeling and estimating the extent of reverberation in the audio recording. Liu et al. [6]suggested for future authentication and recovery how to embed compressed signals into the host audio. Yang Cuccovilloet at. [7]proposed to apply the device fingerprint's inconsistency for

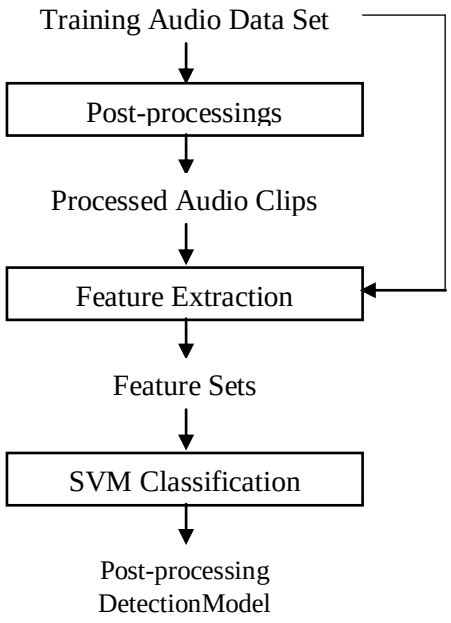
ISSN: 0971-1260
Vol-22-Issue-16-August-2019

tampering detection. Luo et al. [8] proposed to extract audio features from Mel Frequency Cepstral Coefficient (MFCC) and Modified Discrete Cosine Transform (MDCT), based on which the audio waveform compression history can be estimated. Luo et al. [9]suggested Amplitude Co-occurrence Vector (ACV) features that manipulate co-occurrence patterns in audio signals so that original audio signals and post-processed audio signals can be separated.

The Mel-Frequency Cepstrum (MFC) in audio signal processing is a portrayal of an audio signal's momentary power spectrum established on a linear cosine transformation of a log power spectrum on a nonlinear frequency Mel scale.The MFCCs are coefficients that in combination with each other constitute an MFC, and one of the most common voice recognition features is the MFCC function. In addition to MFCC, Linear Prediction Coding (LPC)[2]is one of the most popular approaches to modeling the production of human voice.This article, therefore, has reviewed a method for spotting audio post-processing using the host audio clip's MFCC attribute and LPC feature; with the audio traits, the support vector machine (SVM) classifier is used to authenticate originality of host audio and also to identify the specific post-processing operations.

2. Post-processing Detection Algorithm.

The detection of audio post-processing and operation detection method involves two basic steps: training and testing. Figure I. shows the training procedures of audio post-processing detection and Figure IIshows the flowchart of audio post-processing authentication and operation identification. During the training procedures, firstly the trained post-processing is applied into the audio clips of training dataset, thus generating the post-processed audio clips. Then feature extraction is performed into both the training audio clips and post-processed audio clips to extract the corresponding feature sets. Finally, the SVM classifier is applied to train the feature sets to create the model for post processing detection use. After the training procedure, in Figure II, to detect the existence of post-processing in the input audio clips, for the creation of the feature sets, the same extraction method is applied.



ISSN: 0971-1260
Vol-22-Issue-16-August-2019

Figure I. Training procedures for the detection of audio post-processing[3]

By comparing the generated feature sets with the above-mentioned post-processing detection model, the SVM classifier will create a prediction tag that will help to determine the occasions if audio clips are post-processed or not; it will also be possible to detect the different post-processing events of the post-processed audio clips [3].

The MFCCs and LPC which was apply for feature extraction will be explained respectively in the following section. And for further authentication, the demonstration of the trained feature sets is done by applying the SVM method.

A. Mel-Frequency Cepstrum Coefficient (MFCC)

To extract the MFCCs from the input audio clip, as given in FigureIII, firstly divide the audio clip into frames, periodogram spectral estimation of power spectrum in each frame, $P_i(k)$, is calculated using (1), this helps to identify the frequencies in the appropriate frame.

$$P_i(k) = \frac{|S_i(k)|^2}{L_{Frm}} \qquad (1)$$

The corresponding i^{th} frame is estimated by $P_i(k)$ which is the periodogram-based power spectral, wherei represents the frame number; frame length is represented as L_{Frm}, which denotes the number of sample cases in each frame; and Si(k) is the i^{th} frame's complex DFT, calculated using the following. (2).

$$S_i(k) = \sum_{i=1}^{L_{Frm}} s_i(l)h(l)e^{\frac{-j2\pi kl}{L_{Frm}}} \qquad 1 \le k \le K \quad (2)$$

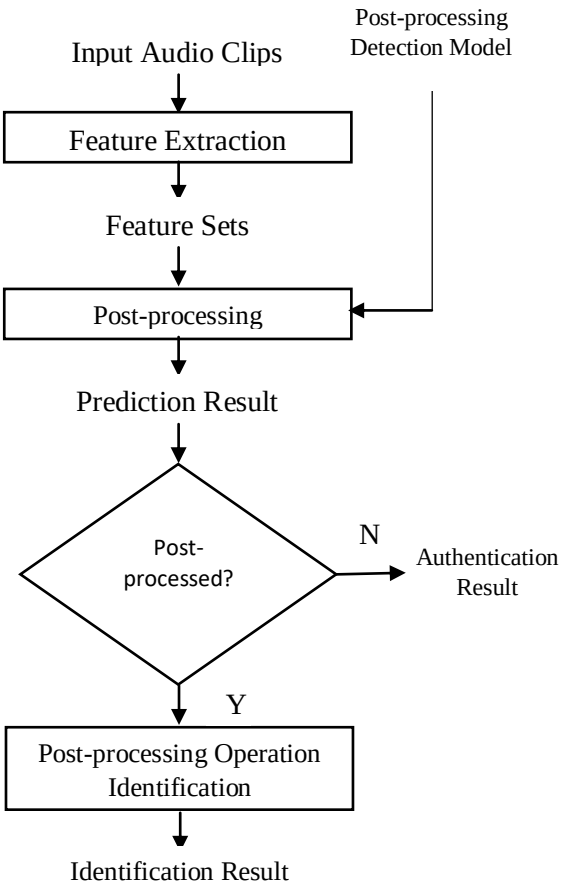
Where h(l) is a long hamming window of the L_{Frm} sample; l indicates the number of the sample in the respective frame; and K is the DFT length.

Then the mel-spaced filterbank, Hm, consisting triangular filters, is determined using (3) and added to the power spectrum for the calculation of filterbank energies by multiplying the filterbank with Pi(k) and then the coefficients were added.

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k \le f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \le k \le f(m+1) \\ 0 & k > f(m+1) \end{cases} \qquad (3)$$

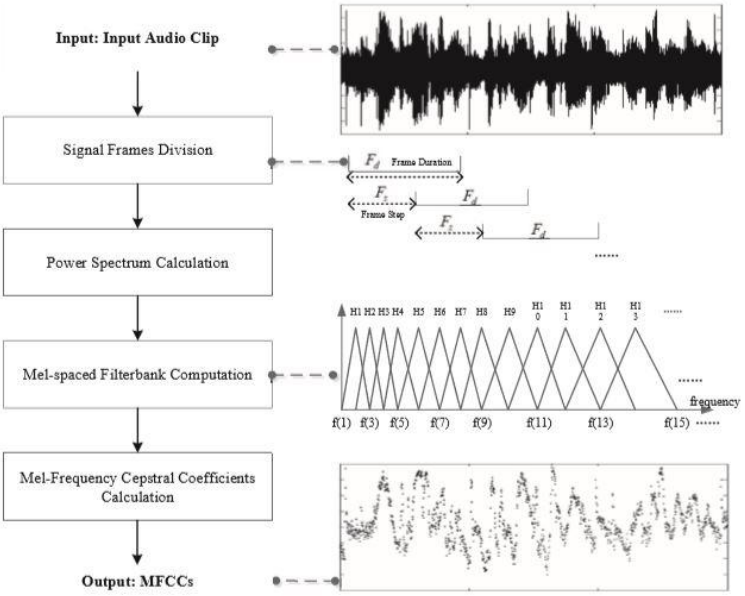
Where m means the number of the filter, and f() denotes various frequencies mel-spaced.

ISSN: 0971-1260
Vol-22-Issue-16-August-2019



FigureII. Flow chart for operation identification and post-processing authentication[3]

Finally, the logarithm of each one of the filerbank energies are computed and the cepstral coefficients are calculated using the DCT. Details of the calculation of the MFCC is stated by Yuan et al.[1].

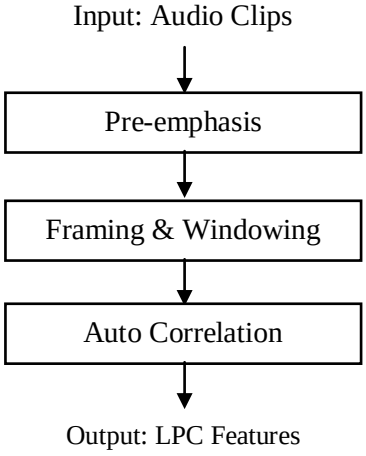


ISSN: 0971-1260
Vol-22-Issue-16-August-2019

Figure III. MFCC's calculation procedure[3].

B. Linear Predictive Coding (LPC)

In addition to the MFCC, as a key feature utilized in auditory and processing of speech signal which draws out various parameters of speech like spectra and pitch formats, LPC is also referred as temporal approach which was primarily invented to make it equivalent to the resonant structure of human vocal tract that developed the subsequent sound. Therefore,the current study hasreviewed the method to apply LPC for audio post-processing detection. Figure. IV shows the procedures of LPC feature extraction. Firstly a filter with coefficient between 0.9 and 1 is applied to the input audio clip to spectrally flatten the audio clip thus making it less susceptible to precision effect. Then the pre-emphasized audio clip will be divided into frames which will be multiplied by hamming window with purpose of minimizing the edge effect. During the framing, to ensure stationary between frames, there is a standard



overlap of 10ms between two adjacent frames. Finally the auto-correlation is applied to calculate the corresponding LPC features[2].

Figure IV. LPC feature extraction procedure[3]

3. Experiments and Results

In the given experiments, the speech audio signal from the GTZAN Speech Dataset [10]has been used to check the efficiency of the reviewedmethod. The GTZAN Speech Dataset compose of 120 audios, each 30 seconds of length and the audios are 22.05 kHz Mono 16-bit audio files, which are in ".wav" format. Considering the time complexity, the audios are divided into shorter clips of 3s length[3].

As introduced in section 2, a training dataset is required to create the model for further post-processing detection use, therefore two datasets were created – 'Training Dataset' and 'Testing Dataset', each of which contains 100 source audio clips. Then, the Adobe Audition (previously COOLEDIT) is applied to perform the post-processing on the two datasets. Four kinds of post-processing operations are performed on the half randomly selected audios from the two datasets respectively. Table I displays the post-processing operations that were applied: Compression, with the compression ratio as 48kbps; High-Pass Filtering, with the

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

frequency set as 1kHz; Band-Pass Filtering, with the bandwidth set as 1800Hz (200Hz to 2000Hz) and Hiss-Reduction[3].

TABLE I. Trained post-processing operations[3]

Serial No.	Operation	Description	Abbr.
I	Compression	Compress a .WAV file into an .MP3 file	COM
Ii	High-Pass Filtering	Audios through a high-pass filter	HPF
Iii	Band-Pass Filtering	Audios through a band-pass filter	BPF
Iv	Hiss-Reduction	To reduce the buzz	HIS

A. Audio Post-processing Detection Results

To judge whether an audio clip from the Testing Dataset was post-processed or not. Firstly, the Training Dataset is performed with the four post-processing operations respectively, thus generating four post-processed Training Datasets; afterwardsthe training procedures were applied, as illustrated in Figure II, to the source Training Dataset and the corresponding post-processed Training Datasets; in this way, the model is created for audio post-processing detection. In this experiment, the frame duration of MFCC is set to be Fd = 1152, and the frame overlap is set to be half of the frame duration, Fc = 576. To find the best parameter, the Mel filter amount P = {16, 24, 32} should vary with three different feature dimensions, 12D, 18D and 24D.

To indicate the post-processing detection result, a label set c={0,1} is defined, where ‘0’ denotes the original audio clip and ‘1’ denotes the post-processed audio clip. Accuracy of the post-processing detection results when using different Mel filter amount are given in Table II, Table III and Table IV, respectively, which reveal the good performance of the reviewed method[3].

TABLE II. Accuracy for Post-processing Detection with MFCC Feature (P=16) (%)[3]

P=16	12D	18D	24D
COM	99	100	100
HPF	100	100	100
BPF	100	100	100
HIS	100	100	100
Average	99.75	100	100

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

TABLE III. Accuracy for Post-processing Detection with MFCC Feature (P=24) (%)[3]

P=24	12D	18D	24D
COM	99	100	100
HPF	100	100	100
BPF	100	100	100
HIS	92	91	92
Average	97.75	97.75	98

TABLE IV. Accuracy for Post-processing Detection with MFCC Feature (P=32) (%)[3]

P=32	12D	18D	24D
COM	100	46	47
HPF	100	100	100
BPF	100	100	100
HIS	100	91	92
Average	100	86.5	86.75

Similarly, as using MFCC, the LPC feature is applied. As shown in Table V, the accuracy of post-processing detection is evaluated when LPC orders vary from 8 to 16. The superior results stipulate the high performance of the reviewedmethodology.

TABLE V. Accuracy for Post-processing Detection with LPC Feature (8-16) (%)[3]

Order	8	9	10	11	12	13	14	15	16
COM	76	79	88	93	96	94	98	97	97
HPF	96	95	99	95	100	100	100	100	100
BPF	100	100	100	100	100	100	100	100	100
HIS	97	94	98	100	100	100	100	99	99
Average	92.25	92	96.3	97	99	98.5	98.5	99	99

B. Audio Post-processing Operations Identification Results

Besides simply judging whether the audio clip has been post-processed or not, this review can furthermore distinguish the specific post-processing operation if the audio clip is claimed to a post-processed one. Considering the time complexity, 12-dimension MFCCs is used when the Mel filter amount P=32 will produce the highest post-processing detection accuracy; therefore, 12D and P=32 are selected for the consequent post-processing operations identification when using MFCC[3].

Table VI shows the confusion matrix for post-processing operation identification using MFCC feature, where the diagonal results indicate accuracy of the

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

corresponding post-processing identifications. The results indicate that the average accuracy of post-processing operation identification using MFCC can be up to 97.4%.

TABLE VI. Confusion Matrix of Post-Processing Operation identification Using MFCC (Mel Filter =32, 12d) (%)[3]

	ORI	COM	HPF	BPF	HIS
ORI	100	7	*	*	*
COM	*	87	*	*	*
HPF	*	*	100	*	*
BPF	*	*	*	100	*
HIS	*	6	*	*	100

a. The average accuracy of the main diagonal is 97.40%
b. Symbol '*' denotes the value is 0%

TABLE VII. Confusion Matrix of Post-Processing Operation Identification With LPC (Order=12) (%)[3]

	ORI	COM	HPF	BPF	HIS
ORI	91	8	*	*	*
COM	3	91	*	*	*
HPF	*	*	100	*	*
BPF	*	*	*	100	*
HIS	6	1	*	*	100

a. The average accuracy of the main diagonal is 96.40%
a. Symbol '*' denotes the value is 0%

Same as in post-processing detection, the LPC feature is applied for post-processing operation identification as well. Table VII shows the confusion matrix for post-processing operation identification with LPC feature, where the diagonal results indicate accuracy of the corresponding post-processing identifications. The results indicate that the average accuracy of post-processing operation identification using LPC feature can be up to 96.4%.

C. Comparisons

The reviewed method has been compared with the already existing method [9] in both post-processing detection and identification of operation, as shown in Table VIII. For fair comparison, the medium results of post-processing detection were selected, when the 12 dimensions and 24 Mel filters are applied with MFCC feature and the order = 12 is selected with LPC feature, instead of the highest accuracy which is achieved. The comparison results indicate the high performance of the working methodology.

TABLE VIII. Comparison of Reviewed Method and Existing Method (%)

	ACV[9]	MFCC [3]	LPC [3]

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

Post-processing detection	95.41	97.75 (12D, P=24)	99 (Order=12)
Operation Identification	94.59	97.40 (12D, P=24)	96.40 (Order=12)

4. Conclusions

In this study, a method for the detection of audio post-processing and operation identification has been reviewed on the basis of features of MFCC and LPC. For learning and classification, the SVM is asked for.Experimental findings suggest that both post-processing detection and procedure recognition are significantly affected by the reviewed method. In future, more post-processing operations will be evaluated and the improvement on audio features will be studied.

References

[1] X. C. Yuan, C. M. Pun, and C. L. Philip Chen, "Robust Mel-Frequency Cepstral coefficients feature detection and dual-tree complex wavelet transform for digital audio watermarking," *Inf. Sci. (Ny).*, 2015.

[2] H. Gupta and D. Gupta, "LPC and LPCC method of feature extraction in Speech Recognition System," in *Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016*, 2016.

[3] Y. Zhan and X. Yuan, "Audio post-processing detection and identification based on audio features," in *2017 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, 2017, vol. 7, pp. 154–158.

[4] P. M. G. I. Reis, J. P. C. L. Da Costa, R. K. Miranda, and G. Del Galdo, "ESPRIT-Hilbert-Based Audio Tampering Detection with SVM Classifier for Forensic Analysis via Electrical Network Frequency," *IEEE Trans. Inf. Forensics Secur.*, 2017.

[5] H. Malik and H. Farid, "Audio forensics from acoustic reverberation," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 1710–1713, 2010.

[6] Z. Liu, F. Zhang, J. Wang, H. Wang, and J. Huang, "Authentication and recovery algorithm for speech signal based on digital watermarking," *Signal Processing*, 2016.

[7] L. Cuccovillo, S. Mann, M. Tagliasacchi, and P. Aichroth, "Audio tampering detection via microphone classification," in *2013 IEEE International Workshop on Multimedia Signal Processing, MMSP 2013*, 2013.

[8] D. Luo, W. Luo, R. Yang, and J. Huang, "Identifying Compression History of Wave Audio and Its Applications," *ACM Trans. Multimed. Comput. Commun. Appl.*, 2014.

[9] D. Luo, M. Sun, and J. Huang, "Audio postprocessing detection based on amplitude cooccurrence vector feature," *IEEE Signal Process. Lett.*, 2016.

[10] G. Tzanetakis, "Music analysis, retrieval and synthesis of audio signals MARSYAS," 2009.

[11] Lakhwani Kamlesh, P D Murarka Arya, and Narendra Singh Chauhan.2015. "Color Space Transformation for Visual Enhancement of Noisy Color Image."International Journal of ICT and Management 2 (October): 2026–6839.