❐   2211

Statistical Based Audio Forensic on Identical Microphones

Fajri Kurniawan[1], Mohd. Shafry Mohd. Rahim[2], Mohammed S. Khalil[3], Muhammad Khurram Khan[4]

[1,2] Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Malaysia
[1,3,4] Center of Excellence in Information Assurance, King Saud University, Riyadh, Saudi Arabia

Article Info

ABSTRACT

Microphone forensics has become a challenging field due to the proliferation of recording devices and explosion in video/audio recording. Video or audio recording helps a criminal investigator to analyze the scene and to collect evidences. In this regards, a robust method is required to assure the originality of some recordings. In this paper, we focus on digital audio forensics and study how to identify the microphone model. Defining microphone model will allow the investigators to conclude integrity of some recordings. We perform statistical analysis on the recording that is collected from two microphones of the same model. Experimental results and analysis indicate that the signal of sound recording of identical microphone is not exactly same and the difference is up to 1% - 3%.

Corresponding Author:

Mohd. Shafry Mohd. Rahim,
Faculty of Computing,
Universiti Teknologi Malaysia,
Skudai, Johor, Malaysia 81310.
Email: shafry@utm.my

1.   INTRODUCTION

Microphone forensics is a recent research interest under audio forensic science. The objective is to authenticate whether a digital audio was made on a given recorder or it has been tampered. Copyright infringement has been an important issue in this 21st century. Identify the microphone model of digital audio recording will provide valuable evidence for the actual ownership when copyright dispute occurs. In addition, forgery on digital audio content is unavoidable nowadays. Using a sophisticated multimedia software make such forgery truly effortlessness. Criminal evidence from arbitrary digital recording must be verified to assure its integrity and originality. Hence, a robust and fast method to authenticate such digital content is progressively vital these days. In addition, information about the recording source can effectively assist other study like gunshot characterization [1], tampering detection [2], speaker recognition [3],[4] and speech enhancement [5].

In general, microphone forensics is a study on the basis of the digital traces that leaves on arbitrary recording. Such traces were occurs due to intrinsic characteristics of the device, which can be shaped from audio sensors, component technology or some defect from the manufacture itself. An audio recording that pretends to have inconsistent traces then it indicated the digital content has been tampered. Actually, tampering audio detection is mostly inspired from previous works in image tampering detection [6].

The pioneers in this field is Kraetzer et al. [7]. Initially, they utilized K-means and Naive Bayes as classifier along with steganalysis features to identify the microphone model. After that, Kraetzer et al. [8] fused Decision Tree and Linear Logistic Regression model to achieve better performance. Few years back, Kraetzer et al [9] proposed huge context model as a guidance for other researcher to select suitable classifier and feature prior to identify the microphone model. The features that commonly used for digital audio data is mel-scaled cepstral coefficients (MFCCs) features. As instance, Brew et al [10] presented speaker verification technique with MFCC as the main features. Dhanalakshmi et al [11] reported digital audio can be

classified to allow better content management. The MFCC feature are fused with linear prediction cepstrum coefficients (LPCCs), perceptually based linear predictive coefficients (PLPCs). Rabaoui et al [12] introduced sound recognition method for surveillance application. They combined three features to achived recognition rate 96.8%. The features are MFCC, wavelet-based and temporal-frequency features. Hanilci and Kinnunen [13] proposed cell-phone recognition from recorded audio that contains speech signal by extracting two features consist of MFCC and LFCC. Eskidere and Karatutlu [14] works on source identification of microphone using multitaper-MFCC features.

In addition, other works done by Buchholz et al. [15]. They considered Fourier coefficient histogram as the features and they utilizing four classifier including Simple Logistic, J48 Decision Tree, KNN and SVM to determine the model. Besides that, Espy-Wilson [16] studied the device identification for landline telephone and microphone. They considered two features named MFCC and linear-cepstral coefficients. The most recent work was performed by Vu et al. [17] in 2012. Vu et al. [17] introduced novel approach called One-class classification (OCC) along with representative instance classification framework (RICF) for microphone forensics. The RICF is introduced to reduce the noisy signal such that it improves OCC performance. Hanilci et al. [18] explored microphone identification problem of speech recording from a mobile phone where audio recorded from 14 models of mobile phones are classified using vector quantization and SVM-based classifier. Moreover, Eskidere [19] reported his recent work in microphone identification on 16 microphone models using GMM-based modeling technique along with three different features called LPCC, PLPC and MFCC.

As presented above, there is only limited studies have been done in microphone forensics. Even vast attempt has been made but current works only considered single device for each various microphone models. In other word, most of them are focus on inter-class problem, which is how to classify some recording then identify the device model among different microphone models. Unfortunately, microphone forensics on identical model of recording device (intra-class problem) is still lack of attention from the community.

In this paper, an audio recording from two identical microphones of the same models are examined using statistical analysis. This study would add new knowledge for microphone forensics community and further stimulate improvement on performance of microphone identification. Wherein, microphone forensics practitioners should not consider only digital traces between one model to another model, but also must take into account the digital traces within the same model in order to carefully assess suspicous digital audio.

In section 1 presents the literature review on microphone forensics and the research motivation. Section 2 presents the specifications of the devices and the explanation of how the audio sample was collected. In Section 3, a brief description of the statistical analysis technique that used in this study. Afterward, the result is discussed and analyzed for each environment. Finally, a conclusion of this study is presented in Section 5 as well as some suggestions for future works.

2.   DATA COLLECTION

The aim of this study is to explore the digital traces within the same microphone model. Hence we collected digital audio recording using two identical microphones of different models. The microphone model that is used in this study is Shure SM-58. There are two identical microphones for that model. It is exciting to study digital traces from more microphones of identical models. However, at this stage we find at least two microphones are enough to explore the difference between identical microphone models if exists. The specification of the microphone that utilized in this study is presented in Table 1.

These two microphones were utilized to collect audio signals at two different locations including quite room and computer laboratory. The quite room is a sound proof room with almost free from noise. On the other hand, computer laboratory introduces many noises from the active CPUs, air conditioner, walking persons and other noises. We considered these two locations to study the frequency response of each microphone against clear and noisy conditions.

Total two recording sessions are performed during data collection. Noted, it is only two sessions because we perform the recording simultaneously for all microphones in each environment. In this regards, all microphones are organized in a row (horizontally) using standard microphone stands. Moreover, one session is a three minutes recording, which consists of both silence and speech recordings. Silence recording means the microphone was passively record the environment without any human speech exists. Meanwhile, speech recording was prepared from a person who reads predefined sentences for two minutes. The person is sitting on a chair and the distance between the speaker and the microphone was fixed to 30 cm. The generated file of audio sample is described in Table 2.

3. STATISTICAL-BASED METHOD

This section briefly describes the statistical analysis technique utilized in this study. We consider five metrics to analyze the signal, i.e. standard deviation, mean, the crest-factor Q, dynamic range D and autocorrelation time. These metrics has been widely known and proven capable to characterize an audio signal [20]. In this experiment, the collected audio signal is analyzed under Matlab environment.

Table 1. Microphone Features And Specifications

Shure SM-58 (Mic$_1$: 2 units)	SPECIFICATIONS
	• Type: Dynamic • Frequency Response: 50-15,000 Hz • Polar Pattern: Cardioid • Sensitivity: -54.5 dBV/Pa • Impedance: 150 Ω (300 Ω actual) • Polarity: Positive pressure on diaphragm produces positive voltage on pin 2 with respect to pin 3 • Connector Type: 3-pin XLR • Net Weight: 298 grams • Dimensions: 162 mm L x 51 mm W

Table 2. The Audio Sample Description

Description	Value
Format	Wave
Audio Format	PCM
Codec ID	1
Bit rate	705.6 Kbps
Channel(s)	1 channel
Sampling rate	44.1 KHz
Bit depth	16 bits
File size	~16.9 MB
Overall bit rate mode	Constant
Bit rate mode	Constant
Format settings, Endianness	Little
Format settings, Sign	Signed

3.1. Mean Value Of A Signal

The average value of the signal or known as the mean (indicated by μ) is used as the basis to measure the signal power. It is also known as the direct current value (DC value). Repetitive signal such as sine wave can be described simply using the DC value. Unfortunately, most of natural signals, e.g. speech, noise or music, normally have random peak-to-peak amplitude. Hence, standard deviation of signal can be utilized to describe such signal. The mean can be calculated as follow $= \frac{1}{N} \sum_{i=0}^{N-1} x_i$, where the signal stored in xi, with number of samples N signal.

3.2. Standard Deviation Of A Signal

The signal fluctuation from its mean and fluctuation power can be estimated from standard deviation of a signal, called σ (sigma). Basically, σ is quite similar with average deviation. The different is the averaging in σ using the power, not the amplitude. Calculation of σ is depicted as below:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \mu)^2$$

Where, the signal defined as xi. Then, Mu is mean of the signal and N is number of samples.

3.3. The Crest-Factor Q

Quantity of impulsive noise, short events or shocks can be estimated using crest factor. It is able to compute the probability of unnecessary wave with respect to the mean of signal. Such unwanted signal can be considered as distortion or bit error of the signal. Crest factor is measuring a waveform's peaks with respect to mean value. Natural sound normally has high crest factor, whereas crest factor equal to one means the signal has no peaks. Crest-factor calculated in base-10 logarithmic form as below:

$$CF = 20 \times log^{10} \left(\frac{V_p}{V_{rms}} \right)$$

Where Vp is peak amplitude of the signal and Vrms is root mean square.

3.4. Dynamic Range D

Dynamic range is simply the ratio between peak and bottom of a signal. In audio signal, dynamic range is commonly computed as base-10 logarithmic value. It often used to express ratio of the loudest possible wave with respect to RMS of noise amplitude. Dynamic range of human speech on average is around 40 dB [20]. Equation below computed the dynamic range of signal:

$$DR = 20 \times log^{10} \left(\frac{V_{peak}}{V_{bottom}} \right)$$

Where, Vpeak and Vbottom are peak and bottom of signals, respectively.

3.5. Autocorrelation Time

Autocorrelation is a function to measure similarity between its original signals versus time-lag applied on same the signal. This function has capability to search repetitive pattern that suppressed with another signal. Autocorrelation commonly applied on statistical signals. Equation below defined the discrete autocorrelation R of discrete signal:

$$R_{yy}(l) = \sum_{n \in Z} y(n) \bar{y}(n-l)$$

In general, the autocorrelation related to a delay time t is determined as follow:
1) Calculate signal value at a time t, denoted as S_1
2) Calculate signal value at a time $t + \tau$, denoted as S_2
3) Computed multiply of two signals, $S_{t-t+\tau} = S_1 \times S_2$
4) Perform steps 1-3 for all desired times t
5) Finally, calculate the average $\bar{S} = \frac{1}{T} \sum_{t=0}^{T-1} S_{t-t+\tau}$.

4.   RESULT AND DISCUSSION

The collected audio contents are analyzed using statistical analysis technique as presented above. Firstly, the audio signal is separated between silence recording and speech recording. The separation is simply based on time frame. The first 60 seconds are considered as silence recording and after 60 seconds are tagged as speech recording. In this experiment, the comparison is taking between identical microphone, same recording type and at same environment. For example, silence recording at quite room from microphone Shure SM-58 is compared with silence recording at quite room of Shure SM-58 of another microphone.

Four graphical plots includes signal in time-domain, amplitude spectrum, histogram, autocorrelation are depicted and compared in a table manner to visualize the difference if exists. In addition, five statistical values are calculated then compared among identical microphones.

4.1. Microphone Forensics In Quite Room Recordings

In the quite room, we expect the microphone should give higher similarity among them. As it is known that no noise was present and the microphone should not capture any noisy signal during silence recording. However, the experiment show an opposite result as expected. Figure 1 shows how signal of silence recording of Mic_{1a} in the time-domain is bigger than Mic_{1b}. Moreover, we found also Mic_{1b} performs anomalous based on observation on autocorrelation plot. The autocorrelation plot explained that the captured silence using Mic_{1b} has several patterns. Meanwhile, Mic_{1a} shows only single pattern exists on the signal, which is make sense for silence signal in quite room. Notice, both microphone are recording simultaneously.

Refer to statistical analysis in Table 3, the big difference between Mic_{1a} and Mic_{1b} is the dynamic range value. Mic_{1b} produce higher dynamic range rather than Mic_{1a}. This means Mic_{1b} generate more noise in silence recording. The reason probably there is manufacture defect on Mic_{1b}. Then, the peak crest factor Q of Mic_{1b} also higher than Mic_{1a}. Finally, both identical microphone produce similar maximum autocorrelation value that near to 55 seconds.

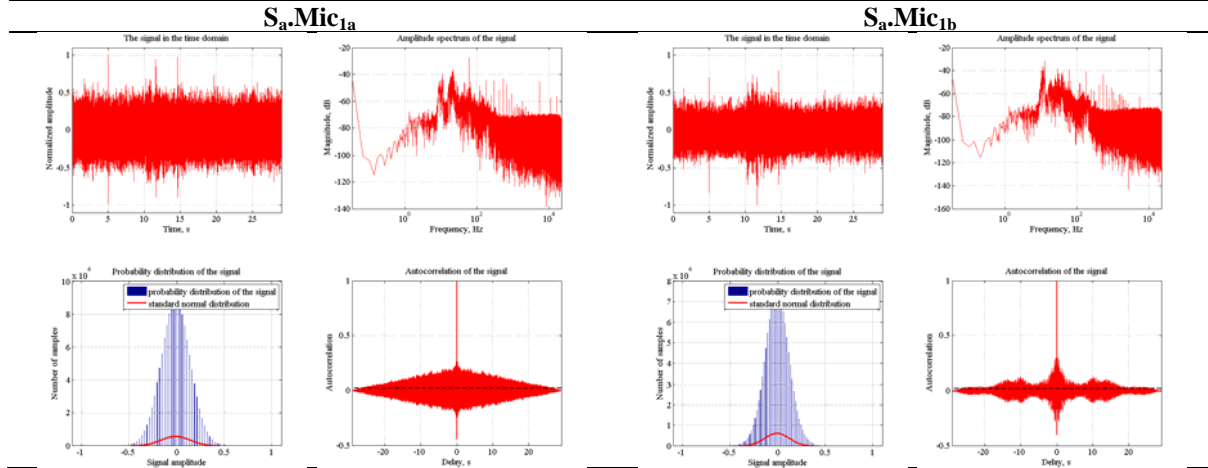| $S_a.Mic_{1a}$		$S_a.Mic_{1b}$	

Figure 1. From top to bottom: Plot for signal in time-domain, amplitude spectrum, probability distribution and autocorrelation of the silence recording in quite room

Table 3. Statistical Analysis of Silence Recording in Quite Room For Shure SM-58 (Mic_{1a} and Mic_{1b})

Metrics	$S_a.Mic_{1a}$	$S_a.Mic_{1b}$	$\mid S_a.Mic_{1a} - S_a.Mic_{1b} \mid$
Sigma	0.15204	0.12433	0.027710
Mu	-0.014822	-0.010417	0.004405
Peak (crest) factor Q (dB)	16.3198	18.0784	1.758600
Dynamic range D (dB)	31.5957	34.6479	3.052200
Autocorrelation time (sec.)	54.9212	54.9202	0.001000
Average Difference			0.968783

Experiment result for speech recording shows more stable with high similarity compare to silence recording as explained previously. Figure 2 described the similar signal, amplitude, histogram and autocorrelation produces by both microphones. The statistical analysis as presented in table 4 denoted the average difference of the two identical images is much smaller compare to the silence recording in Table 3.

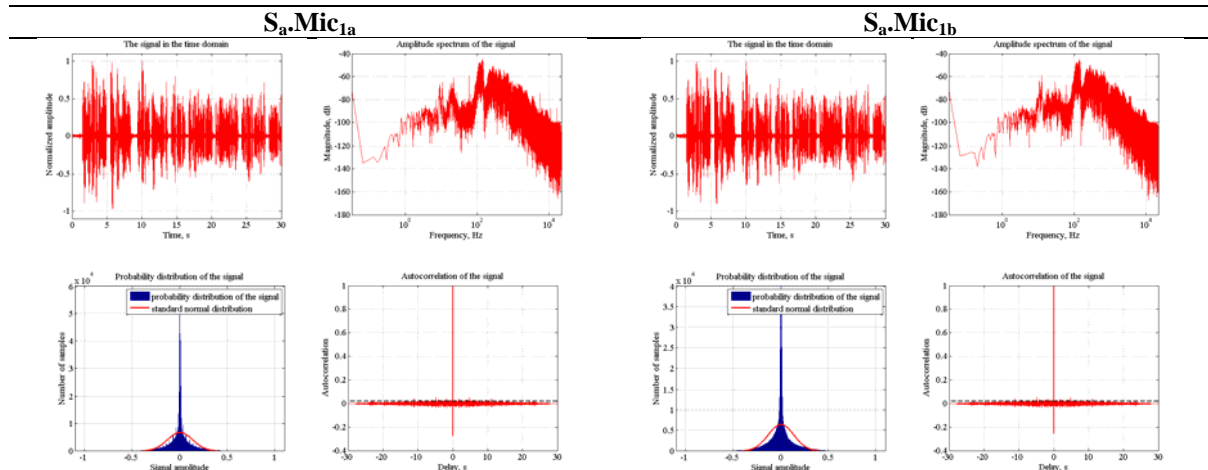| $S_a.Mic_{1a}$		$S_a.Mic_{1b}$	

Figure 2. From top to bottom: Signal in time-domain, amplitude spectrum, probability distribution and autocorrelation of the speech recording in quite room

Table 4. Statistical Analysis of Speech Recording in Quite Room For Shure SM-58 (**Mic$_{1a}$** and **Mic$_{1b}$**)

Metrics	S$_a$.Mic$_{1a}$	S$_a$.Mic$_{1b}$	| S$_a$.Mic$_{1a}$ - S$_a$.Mic$_{1b}$ |
Sigma	0.10204	0.10667	0.00463000
Mu	-0.00040956	-0.00042485	0.00001529
Peak (crest) factor Q (dB)	19.8242	19.4387	0.38550000
Dynamic range D (dB)	62.7913	62.4836	0.30770000
Autocorrelation time (sec.)	0.091293	0.07068	0.02061300
Average Difference			0.14369166

4.2. Microphone Forensics In Computer Lab Recordings

In lab recordings, Mic$_{1b}$ again shows strange autocorrelation on silence recording as presented in Figure 3 and the difference on autocorrelation value is higher than in quite room (Table 3). Meanwhile, another metrics (sigma, mu, peak crest-factor, dynamic range) does not show much difference compare to Mic$_{1a}$, as shown in Table 5.

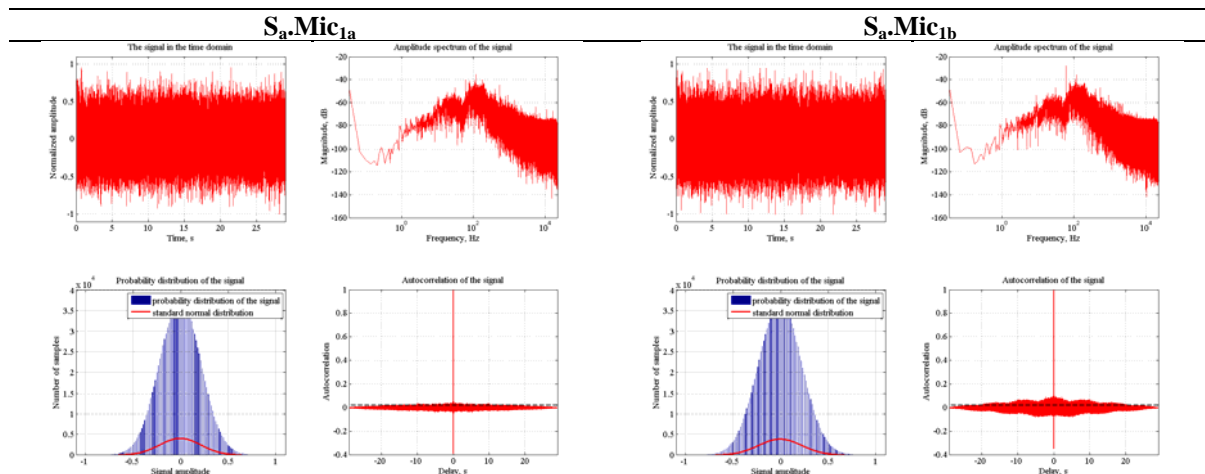| S$_a$.Mic$_{1a}$	S$_a$.Mic$_{1b}$

Figure 3. From top to bottom: Signal in time-domain, amplitude spectrum, probability distribution and autocorrelation of the silence recording in computer lab

Table 5. Statistical Analysis of Silence Recording in Computer Lab. For Shure SM-58 (**Mic$_{1a}$** and **Mic$_{1b}$**)

Metrics	S$_a$.Mic$_{1a}$	S$_a$.Mic$_{1b}$	| S$_a$.Mic$_{1a}$ - S$_a$.Mic$_{1b}$ |
Sigma	0.21418	0.20988	0.004300
Mu	-0.0088191	-0.0085453	0.0002738
Peak (crest) factor Q (dB)	13.3772	13.5534	0.176200
Dynamic range D (dB)	36.1236	36.3909	0.267300
Autocorrelation time (sec.)	31.2285	45.2657	14.037200
Average Difference			2.89705476

Looking into the speech recording, we found both identical microphones produce almost the same values. Figure 4 depicts plots for speech recording in computer laboratory. Moreover, from Table 6, it can be concluded that in this speech recording both identical microphones are producing quite similar signals. In addition, comparing this result with previous result for speech recording in quite room (Figure 2 and Table 4), both results shows almost similar pattern. In another sense, speeches recording in the quite room and the computer laboratory have similar statistical values.
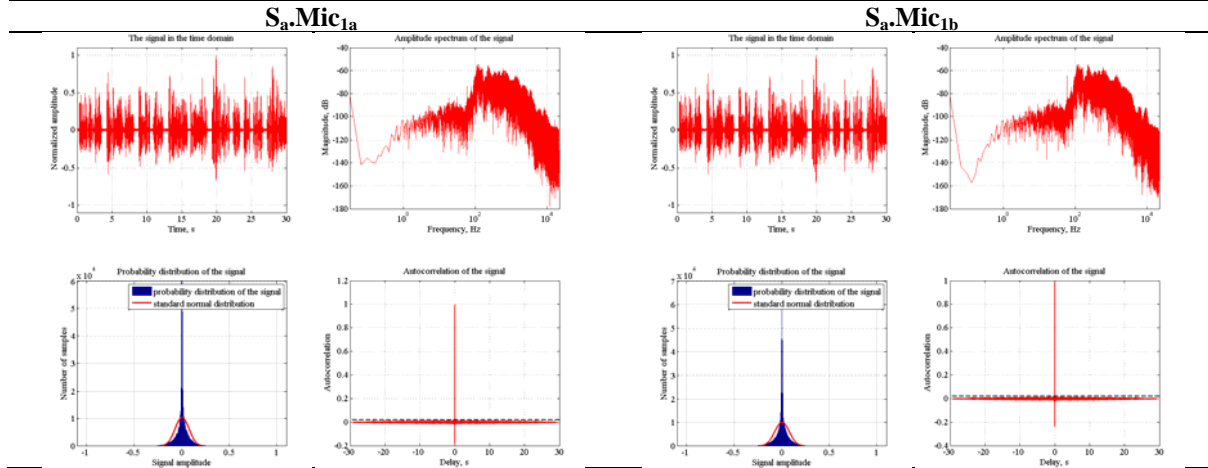
$S_a.Mic_{1a}$	$S_a.Mic_{1b}$

Figure 4. From top to bottom: Signal in time-domain, amplitude spectrum, probability distribution and autocorrelation of the speech rec. in comp. lab

Table 6. Statistical Analysis of Speech Recording in Computer Lab. For Shure SM-58 (**Mic$_{1a}$** and **Mic$_{1b}$**)

Metrics	$S_a.Mic_{1a}$	$S_a.Mic_{1b}$	| $S_a.Mic_{1a}$ - $S_a.Mic_{1b}$ |
Sigma	0.06243	0.066088	0.00365800
Mu	-0.00012704	-0.00012876	0.00000172
Peak (crest) factor Q (dB)	24.0922	23.5976	0.49460000
Dynamic range D (dB)	72.93	72.8096	0.12040000
Autocorrelation time (sec.)	0.071315	0.07127	0.00004500
Average Difference			0.12374094

5. CONCLUSION

In this paper, we have presented a novel technique to identify the microphone model. Microphone classification and verification are crucial tasks to ensure originality of some information, which are grow to be more and more important these days in unlawful investigation. As conclusion, our works prove that digital traces not only present in different microphone models but also it found in identical model. The plot of signal in time-domain, amplitude, histogram and autocorrelation are presented to analyze the difference between generated audio signals. Moreover, five statistical values are computed from each signal and used as comparison tools. The experimental result proven that digital traces on identical microphones are different up to 1% - 3%. Hence, forensic expert should consider this difference prior to analyze the integrity of audio content.

In addition, this work can be used as a base to improve the technique for microphone identification in larger space of microphone models. Further works can explore more number of identical microphones and models. Various statistical analysis technique that capable to characterize the audio signal can be considered to study more the difference between identical microphones. Common features as reported in the literature, e.g. MFCCs, LPCCs, etc., that used for microphone identification have to examine again by including audio recording of identical microphones in the testing dataset.

REFERENCES

[1] I. L. Freire and J. A. Apolinario, "Gunshot detection in noisy environments," in *Proceeding of the 7th International Telecommunications Symposium, Manaus, Brazil*, September, 2010.
[2] L. Cuccovillo, *et al.*, "Audio tampering detection via microphone classification," *Multimedia Signal Processing (MMSP), 2013 IEEE 15th Int. Workshop on*, Sept. 30 2013-Oct. 2, pp. 177-182, 2013.
[3] L. Burget, *et al.*, "Analysis of feature extraction and channel compensation in a GMM speaker recognition system," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, pp. 1979-1986, 2007.
[4] D. A. Reynolds, "HTIMIT and LLHDB: speech corpora for the study of handset transducer effects," *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1535-1538, 1997.
[5] Y. Jiang, *et al.*, "A Dual Microphone Speech Enhancement Algorithm for CloseTalk System," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol/issue: 12(6), pp. 4475-4484, 2014.
[6] H. Jin, "Research of Blind Forensics Algorithm on Digital Image Tampering," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol/issue: 12(7), pp. 5399-5407, 2014.

[7]　C. Kraetzer, *et al.*, "Digital Audio Forensics: A First Practical Evaluation on Microphone and Environment Classification," in *Proceedings of the 9th workshop on Multimedia and security, Dallas, Texas*, pp. 63–74, 2007.

[8]　C. Kraetzer, *et al.*, "Unweighted fusion in microphone forensics using a decision tree and linear logistic regression models," in *11th workshop on Multimedia and security. Princeton, New Jersey, USA: ACM Press*, pp. 49–56, 2009.

[9]　C. Kraetzer, *et al.*, "A context model for microphone forensics and its application in evaluations," *Media Watermarking, Security, and Forensics III*, vol. 7880, 2011.

[10]　A. Brew, *et al.*, "An evaluation of one-class classification techniques for speaker verification," *Artificial Intelligence Review*, vol/issue: 27(4), pp. 295–307, 2008.

[11]　P. Dhanalakshmi, *et al.*, "Classification of audio signals using ANN and GMM," *Applied Soft Computing*, vol/issue: 11(1), pp. 716–723, 2011.

[12]　A. Rabaoui, *et al.*, "Using one-class svms and wavelets for audio surveillance," *IEEE Transactions on Information Forensics and Security*, vol/issue: 3(4), pp. 763–775, 2008.

[13]　C. Hanilçi and T. Kinnunen, "Source Cell-Phone Recognition from Recorded Speech Using Non-Speech Segments," *Digital Signal Processing*, vol. 35, pp. 75-85, 2014.

[14]　Ö. Eskidere and A. Karatutlu, "Source microphone identification using multitaper MFCC features," *9th International Conference on Electrical and Electronics Engineering (ELECO), Bursa*, pp. 227-231, 2015.

[15]　R. Buchholz, *et al.*, "Microphone Classification Using Fourier Coefficients," *LNCS.*, vol. 5806, pp. 235–246, 2009.

[16]　D. G. Romero and C. Y. E. Wilson, "Automatic acquisition device identification from speech recordings," in *Proceeding of IEEE International Conference on Acoustics Speech and Signal Processing, Dallas, Texas, US*, pp. 1806–1809, 2010.

[17]　H. Q. Vu, *et al.*, "Identifying Microphone from Noisy Recordings by Using Representative Instance One Class-Classification Approach," *Journal of Networks*, vol/issue: 7(6), pp. 908-917, 2012.

[18]　C. Hanilci, *et al.*, "Recognition of brand and models of cell-phones from recorded speech signals," *IEEE Trans. on Information Forensics and Security*, vol. 7, pp. 625-634, 2012.

[19]　O. Eskidere, "Source microphone identification from speech recordings based on a Gaussian mixture model," *Turkish Journal of Electrical Eng. & Comp. Sciences*, vol. 22, pp. 754-767, 2014.

[20]　J. Eargle, "Handbook of Recording Engineering," Springer, pp. 4, 2005, ISBN 0-387-28470-2.