Audio Engineering Society

Conference Paper

Presented at the Conference on
Audio Forensics
2019 June 18 – 20, Porto, Portugal

Exploiting Frequency Response for the identification of Microphone using Artificial Neural Networks

Azeem Hafeez[1], Khalid Mahmood Malik[2], and Hafiz Malik[3]

[1]*Department of Electrical and Computer Engineering, University of Michigan, MI, USA*
[2]*Department of computer Science Engineering, Oakland University, MI, USA*
[3]*Department of Electrical and Computer Engineering, University of Michigan, MI, USA*

Correspondence should be addressed to Azeem Hafeez (azeemh@umich.edu)

ABSTRACT

Microphone identification addresses the challenge of identifying the microphone signature from the recorded signal. An audio recording system (consisting of microphone, A/D converter, codec, etc.) leaves its unique traces in the recorded signal. Microphone system can be modeled as a linear time invariant system. The impulse response of this system is convoluted with the audio signal which is recorded using "the" microphone. This paper makes an attempt to identify "the" microphone from the frequency response of the microphone. To estimate the frequency response of a microphone, we employ sine sweep method which is independent of speech characteristics. Sinusoidal signals of increasing frequencies are generated, and subsequently we record the audio of each frequency. Detailed evaluation of sine sweep method shows that the frequency response of each microphone is stable. A neural network based classifier is trained to identify the microphone from recorded signal. Results show that the proposed method achieves microphone identification having 100% accuracy.

1 Introduction

Microphone identification is a process that identifies microphone using its unique artifacts. It can be used as proof of ownership and to authenticate audio recordings [1], [2]. In audio forensic analysis, computing impulse response/frequency response for a microphone is an important task that can be used for microphone identification. The main objective of finding frequency response of the microphone is to classify microphones to link the audio recording to the microphone [3]. Recently, microphone identification became an active research area. Multiple attempts have been made to improve its accuracy. For example, Kraetzer's et al.'s microphone identification algorithm has classified 76% of the microphones accurately [4], Aggarwal et al.'s statistical machine learning based approach reported an efficiency of 90% [5], and Cuccovillo et al.'s open set microphone classification scheme claimed an efficiency of 93% [1]. However, in all these methods impulse response approximation is dependent on the speech characteristics. Essentially, from the recorded signal, these methods estimate clean speech that leads to the estimation of impulse response of the microphone. In estimating the channel response of the microphone, these methods infer the log-spectral magnitude of the channel by subtracting the spectrum from the closest matching template of clean speech average spectrum, in a frame of observed speech [6]. According to best of our knowledge, there doesn't exist any work that estimates the impulse response independent of speech characteristics.

The proposed method analyzes uniqueness of device specific attributes in frequency domain. Sinusoidal audio waves of different frequencies are generated and output against each audio is recorded. The output of audio of each frequency (in frequency domain) is divided by the input (in frequency domain), and in this way frequency response of each audio sinusoidal wave is found and frequency response of microphone is computed. In this study, the environment in which the audio is recorded is kept the same.

In this paper, we propose a method to find the frequency response and classify microphones based upon the frequency response. To validate the proposed approach, six microphones are used in this research study. Out of six, four microphones with labels (M_1, M_2, M_3 and M_4) belong to the one manufacturer with same make and model and the other two with labels (M_5 and M_6) belong to the second manufacturer with same make and model are used to find the impulse response of each microphone. The first model of the microphone selected is $SM58$ and the other is $ST95MK$. Audio for each sinusoidal wave is recorded simultaneously by using *Zoom R*-16 device. Experimental results show that microphones have different impulse response even when all the microphones belong to the same manufacturer and have same make and model.

The rest of the paper is structured as follows. Section-2 explains existing state of the art microphone impulse response estimation from the speech signal. Section-3 outlines the details of experimental settings, dataset used. Experimental results of microphone identification are listed in section-4, it also outlines the details of finding the frequency response of a microphone which is independent of speech signal. Concluding remarks and future directions are discussed in section-5.

2 Related work

In microphone identification, two types of methods are used: informed identification and blind identification. In informed identification [7], [8], there is past information about the microphone characteristics, and by using this information it is recognized whether the given audio belongs to the same microphone or not. The second type of microphone identification is blind identification [1], [2], [6], [9], [10], [11], [12], that doesn't involve any prior knowledge about the microphone. Another type of microphone identification has two categories: intraclass identification and interclass identification. Intraclass identification means that the microphones which are to be identified are of the same make and model belonging to same manufacturer. The second type of microphone identification is interclass in which microphones are of different makes or models. Literature review of microphone identification shows that microphones not only have different characteristics in interclass but also in the intraclass microphone identification. The same idea of identifying transmitter based upon frequency response has also been used in automotive security to detect the transmitter (electronic control unit) based upon frequency response [13].

Detailed review of literature shows that focus of the existing studies remained on the estimation of the impulse response from speech, hence the impulse response is dependent on speech characteristics. Most of the existing methods extract feature set from speech recordings to compute impulse response of underlying microphone. The blind channel estimation method, proposed in [1], is used to compute the feature vector from the input recording. The audio is divided into small frames. For each frame, a feature vector consisting of RASTA-MFCC coefficients is computed. The Gaussian mixture model (GMM)-based modelling is used to estimate the underlying parameters for each coefficient. The estimated GMM parameters π_i, μ_i, $\sum_i$ are used to compute a relative probability matrix P_X which is used to find average log spectrum of the speech [1], [2]. The estimated microphone impulse response can be computed as the difference of average log spectrum of the test file and the estimate of the original clean speech [1].

3 Experimental setup and dataset

Fig.(1) shows the experimental setup used in this research. We find the frequency response of microphones

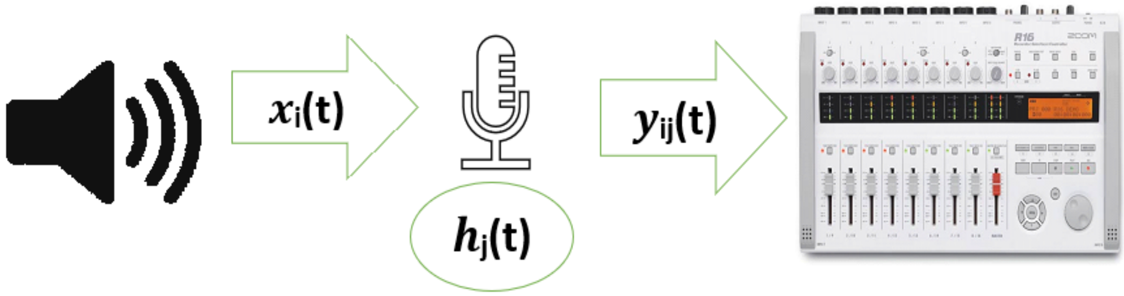

Fig. 1: Experimental setup to determine frequency response

by generating the sinusoidal frequency audio waves, and recording the audio waves by microphone whose frequency response is to be estimated. The frequency of the sinusoidal wave form is increased in equal steps and response to each frequency is recorded using a microphone. Details of dataset used, experimental setup, and experimental results are provided in the following subsections.

3.1 Dataset

The frequency response of the proposed scheme is evaluated using a dataset consisting of 80 frequencies for each microphone. The first audio recording is a sinusoidal wave form of frequency 100Hz, the next audio is of 200Hz, hence with an increment of 100Hz, 80 audio frequencies are generated till 8000Hz. The recording of each frequency tone was of 50 seconds duration. Hence the total duration of signal to compute impulse response for one microphone recording is 4,000 seconds. The impulse response was computed 40 times for each microphone.

3.2 Experimental Setup

Four *SM*58 microphones (same make and model) - labeled as M_1, M_2, M_3 and M_4 and two microphones *ST*95*MK* (same make and model)- labeled as M_5 and M_6– are used to capture audio waves using *Zoom R*-16 digital audio recorder simultaneously in the same acoustic environment. Table-1 shows the details of microphone used in our research. All recordings are made in a small office with surface area covered predominantly with carpet and dry wall. For the dataset collection, the *Zoom R*-16 recorder was set to 44.1 kHz sampling rate. The resolution used was 16 bits/sample.

With the assumption that the microphone behaves like a linear time invariant system, the output for each audio signal is divided by the input signal in frequency domain and the frequency response of the microphone is found.

Table 1: Technical specifications of channels

Label	Manufacturer	Model
M_1	*Shure*	*SM58*
M_2	*Shure*	*SM58*
M_3	*Shure*	*SM58*
M_4	*Shure*	*SM58*
M_5	*AudioTechnica*	*ST95MK*
M_6	*AudioTechnica*	*ST95MK*

3.3 Frequency response estimation

Whenever an input signal $x(t)$ passes through a linear time invariant system $h(t)$, the output signal $y(t)$ is the convolution of $x(t)$ with $h(t)$. Let the impulse response of j^{th} microphone be $h_j(t)$ and let $x_i(t)$ be an audio input to the microphone M_j with impulse response $h_j(t)$, where $h_j(t)$ is to be evaluated. The signal $x_i(t)$ travels through the microphone [14], then $y_{ij}(t)$ which is the response of j^{th} microphone to an audio signal $x_i(t)$ can be expressed as,

$$y_{ij}(t) = x_i(t) * h_j(t) \tag{1}$$

Let $H_j(F)$ be the Fourier transform of $h_j(t)$, $Y_{ij}(F)$ be the Fourier transform of $y_{ij}(t)$, $X_i(F)$ be the Fourier transform of $x_i(t)$. The frequency response of the j^{th} microphone can be estimated as,

$$H_j(F) = \frac{Y_{ij}(F)}{X_i(F)} \tag{2}$$

4 Experimental results

4.1 Evaluation of frequency response of Microphones

Fig.(2) shows the frequency response of microphone M_1 in dB scale. The frequency response of the microphone is computed ten times in the same environment. To test the stability of microphone frequency response, the mean and standard deviation of the frequency response are computed. The following figure consists of three statistical quantities measured vs frequency (in Hz). The first statistical quantity is the mean of 10 readings of frequency response. The second statistical quantity in the following figure is the mean plus standard deviation and the third quantity is the mean minus standard deviation.

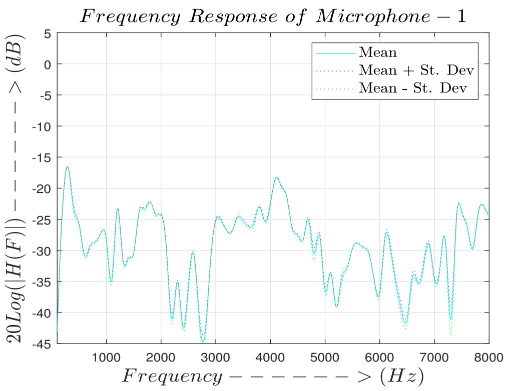

Fig. 2: Frequency response of Mic in dB scale (M_1)

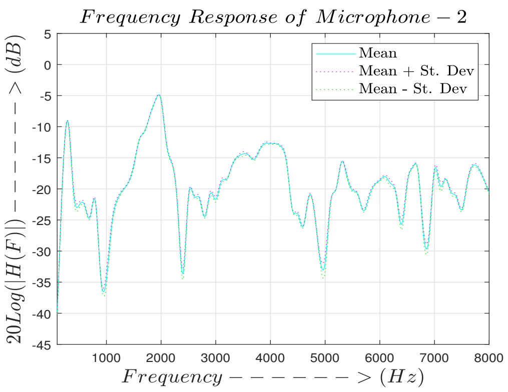

Fig. 3: Frequency response of Mic in dB scale (M_2)

These statistical quantities are computed at all the frequencies starting from 100Hz to 8000Hz with an incremental frequency of 100Hz. From the fig.(2), it is observed that this method of computing frequency is very stable as the mean of the frequency response, mean plus standard deviation and mean minus standard deviation graphs are very close to each other. From this figure, it can be concluded that the frequency response of the microphone is stable. It can also be observed that the microphone frequency response has a very low gain on frequencies around 3,000Hz and high gain on low frequencies and frequency around 4,000Hz.

Fig.(3) shows the frequency response of microphone M_2. The frequency response of microphone M_2 is sketched in dB scale. The figure has three statistical quantities same as for microphone M_1.

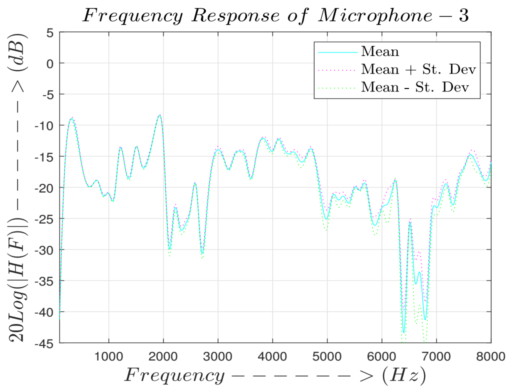

Fig. 4: Frequency response of Mic in dB scale (M_3)

Fig.(4) shows the frequency response of microphone M_3. It is observed that frequency response of M_3 is different than M_1 and M_2.

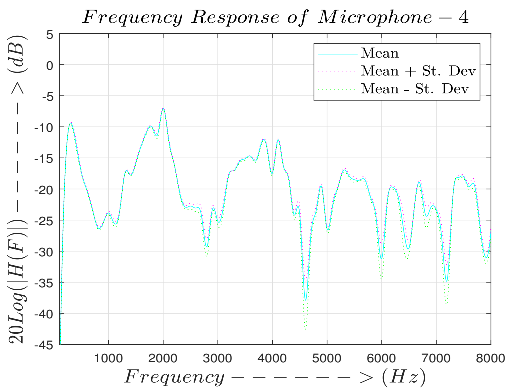

Fig. 5: Frequency response of Mic in dB scale (M_4)

Fig.(5) shows the frequency response of microphone

M_4. Experimental results show that microphones M_1, M_2, M_3 and M_4, which have the same make and model, same manufacturer have different frequency responses. The reason for this is that at microscopic level, each component cannot have the same type and number of atoms. Therefore, the physical characteristics of each microphone is different.

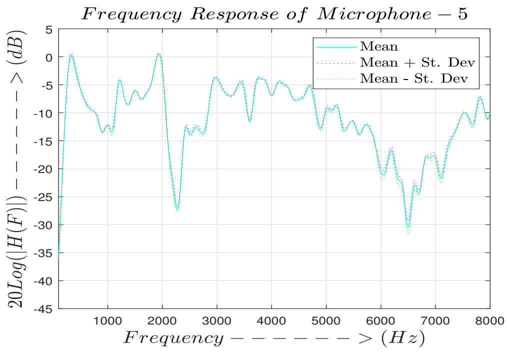

Fig. 6: Frequency response of Mic in dB scale (M_5)

Fig.(6) shows the frequency response of microphone M_5 in dB scale vs frequency. From the fig.(6), it is observed that this method of computing frequency is very stable as the mean of the frequency response, mean plus standard deviation, and mean minus standard deviation graphs are very close to each other. From this figure, it is concluded that the frequency response of the microphone is stable. It is also noticed that the microphone frequency response has a very low gain on frequencies around 2200Hz and 6500Hz. It has high gain on low frequencies and frequency around 1800Hz.

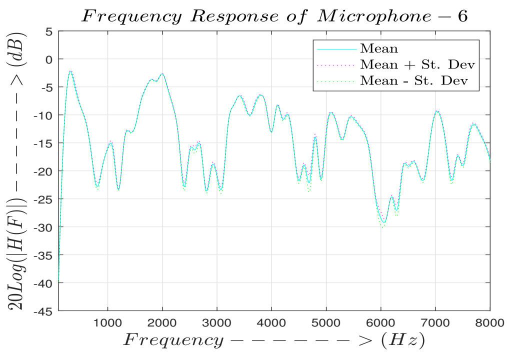

Fig. 7: Frequency Response of Mic in dB scale (M_6)

Fig.(7) shows the frequency response of microphone M_6 on dB scale vs frequency. The frequency response of microphone M_5 and M_6 are different even though both the microphones belong to the same manufacturer and have the same make and model.

4.2 Microphone Identification Using Neural Networks

The purpose of this experiment is to validate that different microphones even from the same make and model introduce different artifacts while transmitting same message over the same channel used. Fig.(8) shows the block diagram of the proposed algorithm used to identify microphones. Audio of different sinusoidal waves is recorded from each microphone and the recording is stored using *Zoom R*-16 device, the frequency response is estimated and stored in the database after classification during training phase. In the testing phase, the frequency response of the test microphone is estimated and trained model is used for classification of each audio file in the testing dataset to identify the microphone. To achieve this goal, the dataset for all the six microphones transmitting the same message is used. In this experiment, the microphone is the only variable and the rest of the variables are kept the same.

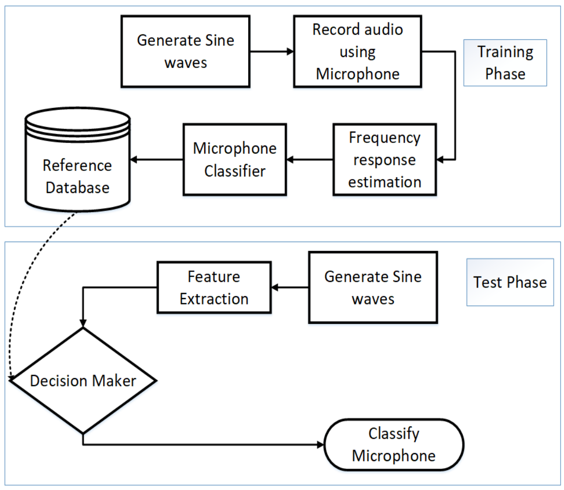

Fig. 8: Microphone Identification using ANN

Scaled conjugate gradient back propagation algorithm is used for microphone classification. Let $p_1, p_2 ...... p_k$ be a set of non-zero vectors weight vectors in $\Re^N$ [15]. The set is said to be conjugate system with respect to non-singular matrix A if the following holds,

$$p_i^T A p_j = 0 \quad (i \neq j, i = 1, 2, ...k) \tag{3}$$

The set of points w in $\Re^N$ satisfying,

$$w = w_1 + \alpha_1 p_1 + \alpha_2 p_2 + ... + \alpha_k p_k, \alpha_i \in \Re \quad (4)$$

where, w_1 is a point in weight space and p_1,p_1......p_k is a subset of conjugate system, is called a k-plane [15]. This algorithm adjusts the weights w_i such that the error is minimized. The iterations continue to find the set of w_i till error is minimized and the local minimum is reached. The main purpose of scaled conjugate gradient back propagation algorithm is to find the weights w_i in the training phase and then weights w_i for classification in testing phase. A multilayer neural network is trained with "scaled conjugate gradient back propagation" training algorithm with 80 input variables (which represents the microphone frequency response) is used. The frequency response is computed 40 times for each microphone.

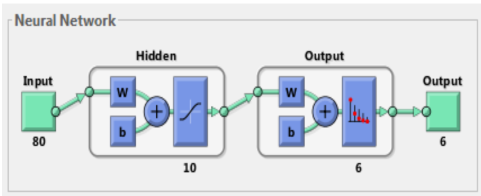

Fig. 9: Neural Network architecture of Microphone classifier

There are 6 outputs from neural network where each output refers to the microphone to which the signal belongs. Stopping criteria of the Epochs = 1000, gradient = 7.5e-7, and one hidden layer with 10 hidden nodes is included. Shown in fig.(9) is the architecture of the multilayer NN trained for channel classification.

Table 2: Confusion matrix for Microphone Classifier

		Target Class						
-	-	M_1	M_2	M_3	M_4	M_5	M_6	Acc. %
Pred. Class	M_1	40	0	0	0	0	0	100
	M_2	0	40	0	0	0	0	100
	M_3	0	0	40	0	0	0	100
	M_4	0	0	0	40	0	0	100
	M_5	0	0	0	0	40	0	100
	M_6	0	0	0	0	0	40	100
	Acc. %	100	100	100	100	100	100	100

Table-2 shows the classification performance of the proposed system in terms of confusion matrix of microphone (M) classification for the training, validation and testing phase. It is observed that the microphone classification achieves an overall detection rate of 100% accuracy in testing, validation and test phase.

5 Conclusion

Microphone identification is an important area in audio forensics. Existing research has focused on the computation of frequency response using the characteristics of speech for microphone identification. Unlike previous studies, this paper presented an approach of using sinusoidal audio waves of different frequencies to compute frequency response. We tested the stability of our method of estimating frequency response. Experimental results showed that different microphones have different gains at different frequencies. Moreover, our results also showed that microphones which belongs to same manufacturer and have same make and model have different frequency responses. our empirical results of also showed that the microphones are identified with 100% accuracy using artificial neural network. In future work, given any speech signal, we will apply inverse filtering on speech signal by using the results of frequency response in this study to identify the microphone.

References

[1] Cuccovillo, L. and Aichroth, P., "Open-set microphone classification via blind channel analysis," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pp. 2074–2078, IEEE, 2016.

[2] Hafeez, A., Malik, H., and Mahmood, K., "Performance of Blind Microphone Recognition Algorithms in the Presence of Anti-Forensic Attacks," in *Audio Engineering Society Conference: 2017 AES International Conference on Audio Forensics*, Audio Engineering Society, 2017.

[3] Hanilci, C., Ertas, F., Ertas, T., and Eskidere, Ö., "Recognition of brand and models of cell-phones from recorded speech signals," *IEEE Transactions on Information Forensics and Security*, 7(2), pp. 625–634, 2012.

[4] Kraetzer, C., Oermann, A., Dittmann, J., and Lang, A., "Digital audio forensics: a first practical evaluation on microphone and environment classification," in *Proceedings of the 9th workshop on Multimedia & security*, pp. 63–74, ACM, 2007.

[5] Aggarwal, R., Singh, S., Roul, A. K., and Khanna, N., "Cellphone identification using noise estimates from recorded audio," in *Communications and Signal Processing (ICCSP), 2014 International Conference on*, pp. 1218–1222, IEEE, 2014.

[6] Gaubitch, N. D., Brookes, M., Naylor, P. A., and Sharma, D., "Single-microphone blind channel identification in speech using spectrum classification," in *Signal Processing Conference, 2011 19th European*, pp. 1748–1751, IEEE, 2011.

[7] Nishimura, A., "Device-Specific Distortion Observed in Portable Devices Available for Recording Device Identification," in *Audio Engineering Society Convention 144*, Audio Engineering Society, 2018.

[8] Uemura, S., Sugiyama, O., Kojima, R., and Nakadai, K., "Outdoor acoustic event identification using sound source separation and deep learning with a quadrotor-embedded microphone array," in *The Abstracts of the international conference on advanced mechatronics: toward evolutionary fusion of IT and mechatronics: ICAM 2015.6*, pp. 329–330, The Japan Society of Mechanical Engineers, 2015.

[9] Thuene, P. and Enzner, G., "Maximum-Likelihood and Maximum-A-Posteriori Perspectives for Blind Channel Identification on Acoustic Sensor Network Data," in *Speech Communication; 13th ITG-Symposium*, pp. 1–5, VDE, 2018.

[10] Makino, S., Lee, T.-W., and Sawada, H., *Blind speech separation*, volume 615, Springer, 2007.

[11] Huang, Y., Benesty, J., and Chen, J., "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, 13(5), pp. 882–895, 2005.

[12] Malik, H. and Zhao, H., "Recording environment identification using acoustic reverberation," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 1833–1836, IEEE, 2012.

[13] Hafeez, A., Tayyab, M., Zolo, C., and Awad, S., "Finger Printing of Engine Control Units by Using Frequency Response for Secure In-Vehicle Communication," in *2018 14th International Computer Engineering Conference (ICENCO)*, pp. 79–83, IEEE, 2018.

[14] Farina, A., "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[15] Møller, M. F., "A scaled conjugate gradient algorithm for fast supervised learning," *Neural networks*, 6(4), pp. 525–533, 1993.