Location Forensics From Media Recordings Using a Novel Pole-matching Classifier

Sayeed Shafayet Chowdhury, *Student Member, IEEE,* Jayanta Dey, *Student Member, IEEE,* Rakib Hyder, *Student Member, IEEE,* Ashraful Islam, *Student Member, IEEE,* and Mohammad Ariful Haque, *Member, IEEE*

***Abstract*—The synthesis as well as the structural, multiferroic and optical characterization of Dy doped $BiFeO_3$ multiferroic materials are presented in this article. Solid state reaction technique was used to synthesize the bulk polycrystalline $Bi_{0.9}Dy_{0.1}FeO_3$ samples, while their nano counterparts were prepared using ultrasonic probe sonication. Significant improvement of phase purity in the as synthesized samples was achieved due to Dy substitution both in bulk $Bi_{0.9}Dy_{0.1}FeO_3$ sample and corresponding nanoparticles which was demonstrated by X-Ray diffraction patterns. Again, FESEM imaging was carried out to analyze the surface morphology of bulk sample as well as to determine the size and distribution of the synthesized nanoparticles. For magnetic characterization, we performed M-H hysteresis loop measurements using SQUID where enhanced magnetic properties were found for Dy doped bulk $BiFeO_3$ compared to their nano samples as well as undoped $BiFeO_3$. Furthermore, leakage current density measurements were performed where, interestingly, we found comparable performance for $Bi_{0.9}Dy_{0.1}FeO_3$ bulk ceramics and its corresponding nanoparticles prepared by using ultrasonic energy. Moreover, higher remanent polarization and coercive field for nanoparticles were observed from the polarization versus electric field measurements compared to their bulk samples which might be associated with leaky behavior generated from oxygen related defects of the nanoparticles. However, intrinsic ferroelectric behavior was observed for .**

***Index Terms*—Location forensics, pole estimation, pole-matching classifier, classification accuracy, computational time.**

I. Introduction

Location forensics has become an important area of research in the 21st century. With the proliferation of terrorism, child pornography [1] or abuse on women, there is an increasing demand for reliable identification of location of media recordings for ease of law enforcing agencies. Success in identifying such locations properly can ease the process of getting hold of the criminals involved. Furthermore, retrieval of location information embedded in the media recordings might pave the way for automatic tagging of geo-location of huge amount of digital data which are being uploaded every moment on social media platforms such as YouTube and Facebook [2]. However, in the absence of concurrent power recordings, this process becomes quite challenging. Again, for security purposes, a consistent as well as fast and computationally light method of region-of-recording recognition is necessary.

A potential route to obtain the information of recording location from media recordings is to extract the fingerprint left on that recording by the power grid of that location. Due to electromagnetic radiation every grid acts like an antenna that radiates an electromagnetic wave containing corresponding grid frequency. It has a nominal value of 50 Hz in North America and 60 Hz in the most other places in the world. This radiation introduces a weak interference in audio or video recordings done near places where there is electrical activity. The instantaneous frequency of a power grid fluctuates over time around its nominal value due to control mechanism and load variation. The variation pattern of the Electric Network Frequency (ENF) over time for any grid is defined as an ENF signal. ENF can vary with time over a certain range for a certain grid. For example, in Lebanon, it varies by about 1 Hz around 50 Hz nominal frequency, whereas in China, this variation is about 0.1 Hz. Though this variation of frequency can be considered a random process, it follows a particular statistical characteristics. The frequency variation is almost identical in all places of the same grid [3] and depends on load variation pattern of that grid. The load variation has different statistical distribution for different grids and follows a specific pattern for a particular grid. So, it can be inferred that the electromagnetic interference contains information about the grid. If such interference pattern can be reliably extracted from the recording, we can use this information to correctly identify the location of the recording. We can detect the location of a media recording by extracting ENF signal from it which is its power signature. As the validity of results depends on the accuracy of the frequency estimates obtained using a relatively short segment of media and sensor recordings, accurate and reliable estimation of ENF signals is of great importance in forensic purposes. The identification of location of media recording is not a trivial task. The recorded signal may contain human speech, acoustic noise, and other interfering signals. The amounts of noise and distortions can be different even within the same signal due to change of recording conditions which can heavily distort the actual ENF signal pattern. A variety of ENF extraction techniques have been reported in the literature. Initial work in this area was proposed by Grigoras and Cooper [4]– [9]. They observed that the ENF is dominated by relatively high energy, sub-0.1 HZ fluctuations caused by generator speed governors keeping the system synchronized. Thus, this dynamics is effectively identical at any location in the power system. This allows a ground truth reference ENF measurement to be obtained from anywhere in the grid which can then be used to determine the time at which the recording was made and to detect tempering. Some of the most recent work that has extended the initial research includes the extraction of ENF from video flicker in [10], refining the Fourier approaches in [11], using dynamic

The authors are with the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh. E-mail: arifulhoque@eee.buet.ac.bd

programming and a feed forward spectral estimator in [12] and incorporating power signal harmonics in [13]. Most of the techniques are based on either time or frequency based methods or variations of these techniques. Both the approaches have practical applications. Time-based techniques such as zero-crossing method have proven to be very useful in order to record the ENF directly from the power line. Such methods are used by power suppliers, which are obliged to keep the ENF within a given tolerance and thus need to record the ENF time history to validate that. Zero crossing method is, however, not suitable to extract ENF from real-world speech or music audio content. Frequency or short-time-Fourier-transform (STFT) based methods, in contrast, are suitable for this, and are commonly applied for this purpose [13]. Again, Garg *et al.* proposed a half-plane intersection method based on highly quantized information from the correlation coefficient between the processed ENF signals across different locations [2]. But the method provides satisfactory results only for power recordings since the audio data do not have high correlation. Additionally, order of estimation constraints increases quadratically with the number of anchor regions [2], making this method quite complex for higher number of locations.

Traditional ENF extraction methods have their own limitations such as ENF segments may be severely distorted due to presence of noise and different extraction method may give different ENF estimations for the same media recording. Traditional ENF based grid detection methods [14] have several stages such as ENF extraction, feature extraction, classification etc. Accuracy of each of these stages depends on the previous stages. This structure may cause loss of valuable information about the grid. In this paper, we propose a novel method to extract features directly from the raw data. In our approach the raw signal is divided into non-overlapping segments of duration of 10 seconds. We estimate the AR parameters from each of the data segments. The roots of the AR-parameter polynomial provide the poles of the system. The roots of different grids are stored that serve as training data set. In testing phase, poles are extracted from the testing data. Now a distance-based classifier is used to match the poles of training data with those the testing data and the grid with minimum average distance between the training and testing poles is estimated as the identified grid. The novelty of our approach lies in the fact that the total classification is performed on features extracted from the raw data itself. Our results demonstrate that the proposed method can successfully classify a given power/audio recording with accuracy of 97.78% on the development set and 94.03% on the test set. The results have been compared with published results based on ENF based classifiers [14], [15] which assert the enhanced performance of the proposed architecture. Moreover, the analysis of the effect of various parameters has been carried out along with a comparative study of the computational overhead between an ENF based classifier and the pole-matching classifier.

The organization of rest of the paper is as follows. Section II defines the problem statement and provides a concise overview on the dataset. Section III presents the factors that restrict the application of ENF based classifiers for the problem at hand. The following section gives a brief account of the poles of a given signal and its significance. The proposed method is described in detail in the next section. Section VI shows the experimental results and comparison with state-of-the-art methods. The paper is concluded with some remarks in Section VII.

TABLE I
LOCATION OF THE GRIDS IN THE DATASET.

Grid Label	Location	Nominal Frequency (Hz)
A	Texas	60
B	Lebanon	50
C	Eastern U.S.	60
D	Turkey	50
E	Ireland	50
F	France	50
G	Tenerife	50
H	India (Agra)	50
I	Western U.S.	60
J	Brazil	60
K	Norway	50
L	Australia	50

II. Data Description and Problem Statement

In this work, the data provided for the IEEE Signal Processing Cup, 2016 [16] have been used. Two types of signals namely audio and power signals are provided for training purpose from 12 different grids and the grids were labeled as A to L. Table I depicts the location of the grids of the dataset along with their corresponding labels. The data consist of 8 grids with nominal frequency of 50 Hz and 4 grids with nominal frequency of 60 Hz. Power recordings were taken by a signal recorder connected to a power outlet using a step-down transformer. However, the audio signal is simply one made using an audio recorder not connected to anything. The given

Power signals are sinusoidal waves with clear periodicity. However, the audio recordings are extremely noisy and thus obtaining the time varying dominant frequency component from such a signal is very challenging. For each grid, 2 audio recordings of duration 30 min each were provided. However, total duration of given power signals of a particular grid was much higher compared to audio signals since several power recording were included in the dataset. This power data were provided for varying amounts of time (5 to 7 hours) for different grids. However, in the development and testing phase, signals were of only 10 min duration. Development data consisted of 45 recordings and testing was done on 134 recordings, from the 12 grids.

The problem is to build an efficient and accurate classifier system which can detect the region-of-recording for any recording. So, the main objective is to first train the system using the available training dataset and ground-truth grid location, then for a test set, locate the grid of recording in a robust and efficient manner.

III. Limitations of ENF based Location Forensics

The state-of-the art methods for region-of-recording classification using media recordings are based on the ENF pattern of the corresponding grid [14]. Figure 1 shows the block diagram of such a method. At the onset, the ENF is extracted

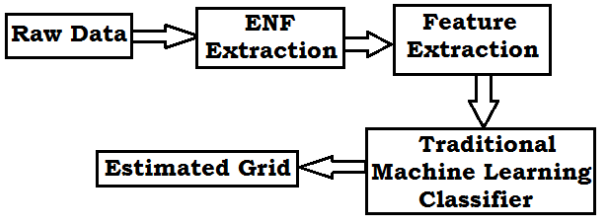
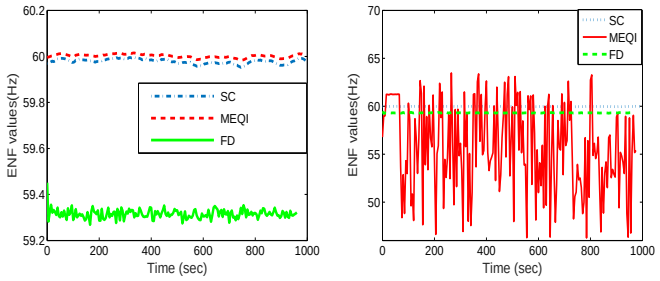


Fig. 1. Block diagram of a traditional ENF based machine learning classifier.



(a) ENF of a power signal (b) ENF of an audio signal

Fig. 2. ENF estimation results on a training data segment of the same signal using different algorithms (SC= spectral combining, MEQI= maximum energy with quadratic interpolation, FD=frequency demodulation).

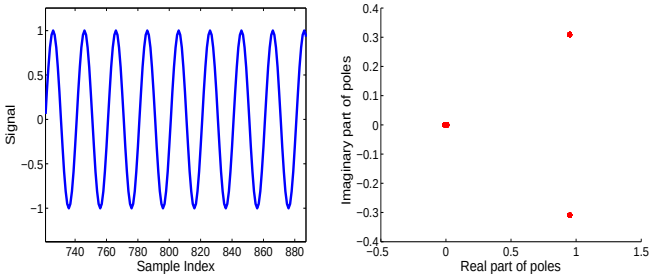
from the power/audio signal and then suitable features are selected from the obtained ENFs to train a machine learning classifier, such as support vector machines (SVM) as reported in [14]. Then the system is trained to classify a given input signal. However, there are some limitations in case of such ENF based location forensics. First of all, since ENF is a power signature of the grid, it should be unique irrespective of the ENF extraction algorithm. But as it turns out, sometimes when the data become quite noisy, this does not hold true. In fact, even for power signals, variations in the obtained ENF pattern are observed as depicted in Fig. 2a. Here, the ENFs extracted are plotted for the same input signal using three different algorithms, spectral combining (SC) [13], maximum energy with quadratic interpolation (MEQI) [17], frequency demodulation (FD) [18]. Since power signals are rather clean, we anticipate the ENFs to overlap mostly. However, as we can see, ENF extracted usign FD differs from the other two greatly. Again, although the patterns of the two ENFs estimated using SC and MEQI are same, they will differ in features such as mean and range of ENF etc. Thus, to some extent, ENF extraction becomes subject to method of estimation and a reliable ENF extraction algorithm becomes pivotal.

On the other hand, for audio recordings, ENF extraction becomes all the more challenging since large noise is present in the signals and obtaining the actual ENF engulfed in such noisy conditions becomes quite difficult. Figure 2b demonstrates an example of this. Since MEQI method [17] seeks to find the ENF based on maximum energy locations from the spectrogram, the ENF fluctuates a great deal in the case of an audio signal where we have much stronger spectral components even outside the nominal frequency and its harmonics. Although spectral combining method [13] tries to obtain the ENF by weighting the harmonics bands depending on the signal-to-noise ratio in the corresponding band, once again extraction of a reliable and unique ENF seems very difficult as ENF patterns from SC and FD methods differ in this case as well.

Additionally, the efficacy of an ENF based location forensic system also depends on feature selection and classifier precision. Since several stages are involved, the inherent grid information embedded in the power or audio signal may be lost if any one of the stages fail to work satisfactorily. As the ENF is not a physically acquired signal, rather an extracted one, the classification accuracy depends on the extraction of ENF greatly. However, if the classification can be made upon the raw data itself, it may preserve some valuable info. This induced us to explore a novel algorithm based on the given raw media recordings and building a classifier based on this which is discussed in the next section.

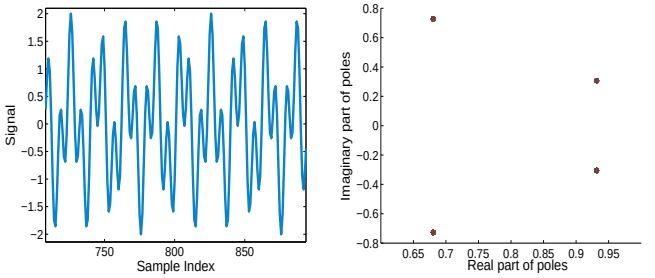
IV. Concept of Pole as a Distinguishing Feature

Having discussed the various problems associated with ENF based classification, this section is dedicated to development of the concept of pole and its relevance to region-of-recording classification. The estimated poles of any given signal represent information about its frequency content in the z-domain. The method of pole estimation is described in a later subsection. As an example, a simple sinusoid and its corresponding pole-positions are plotted in Fig. 3a and 3b, respectively. It can be seen that the estimated poles from the different time windows overlap very closely. As a result, they are grouped together tightly. Again, another example signal is taken as a summation of two sinusoids with 50 and 130 Hz frequencies. A frame of the signal with its respective pole-plot is given in Fig. 4a and 4b, respectively. Once again. the poles are highly localized since the signal contains specific frequencies in it. Interestingly, the frequency content is easily decipherable from



(a) A portion of a 50 Hz sinusoidal signal (b) Pole-plot of the signal in (a)

Fig. 3. A simple sinusoid and its corresponding pole-plot.



(a) A portion of a signal containing a 50 and a 130 Hz sinusoid (b) Pole-plot of the signal in (a)

Fig. 4. A summation signal of two sinusoids and its corresponding pole-plot.

the pole-plots as distinct frequency containing signals result in distinct pole patterns, as evidenced by the above figures. The pole-plots provide a type of signature of the input signal. This triggered our idea of developing a classifier based on the pole-positions of the grid signals. However, the actual instantaneous frequency of a grid will vary over a very small range unlike the ones shown above. Consequently, the pole-plots will not be so highly localized for those cases, rather a more scattered grouping is expected. Although the poles for a grid signal will be spread out over a somewhat wider region, their specific locations and pattern of clustering will vary depending upon the source grid. So, a novel algorithm was developed based on this idea which is explained in the next section.

V. Proposed Method

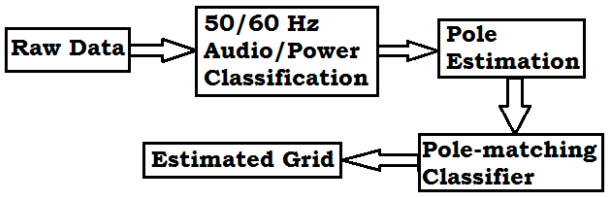
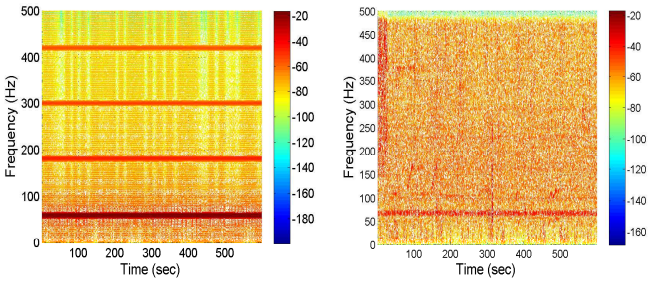


Fig. 5. Block diagram of the proposed pole-matching based region-of-recording classification system.

In this section, we describe the sequence of signal processing tasks carried out to classify a media recording and estimate its grid location. Figure 5 depicts the steps of the proposed scheme. Since we have 4 types of signals, namely 50 or 60 Hz, audio or power, at the very onset we classify the given signal in to one of these 4 categories. Then the poles are estimated from the data and they are stored. In the testing phase, once again, the poles are estimated and then the pole-matching classifier is used to estimate the likelihood of the poles of the testing set belonging to one of the pre-defined grids. The two basic blocks in the proposed structure is pole estimation and pole-matching. In what follows, the proposed method is described in details.

A. Nominal Frequency and Data Type Classification

For any given data, before estimating the poles, it is passed through a nominal frequency and data type classification step. This is done to enhance the inter-grid separability and thus improve the classification performance. Also, since the grids with different nominal frequency can be modeled as the response of different transfer functions, it is not worthwhile to estimate their poles using the same model parameters. Again, this also applies for the audio or power data type as the transfer functions are obviously different. That is why, this sort of classification is done. First the spectrogram of the given signal is computed using short-time Fourier transform (STFT) [19]. The spectrograms of a power and an audio signal are shown in Fig. 6a and Fig. 6b, respectively. As we can see, for the power signal, most of the power lies in the nominal and its odd harmonics. However, for the audio signal, there is noise over the whole spectrum. In fact, the ENF variations becomes engulfed in the noise in this cases since even nominal frequency estimation becomes a challenge. From the Fig. 6b, it seems there is a strong band around 60 Hz, however, this spectrogram is obtained from a signal of a 50 Hz grid. Next,



(a) STFT of a power signal (b) STFT of an audio signal

Fig. 6. Spectrogram estimation using short-time Fourier transform (STFT) of media recordings.

we take the logarithm of the obtained power spectral density (PSD) values. Since we have 2 nominal types 50 or 60 Hz, we then take the summation of the PSD values over these two frequencies and their harmonics, leaving out their common harmonic at 300 Hz. For 50 Hz, we find it as,

$$a = \sum_{c=1, c\neq 6}^{8} \sum_{e=1}^{M} S_{50*c,e} \tag{1}$$

where $S_{x,z}$ is the PSD value at frequency x Hz and z^{th} time instant and we have total M no. of time instants. For 60 Hz,

$$b = \sum_{c=1, c\neq 5}^{8} \sum_{e=1}^{M} S_{60*c,e} \tag{2}$$

If $a > b$, then 50 Hz is chosen as nominal, otherwise 60 Hz is chosen. The reasoning is that most of the power lies in the nominal band and its harmonics. Again, for audio or power type detection, we notice that for the power signal, the nominal and harmonic frequencies occupy maximum amount of power. However, the power is quite spread out over the wide spectrum for the audio signal. So we measure the total power over all frequency bins and time instants and take the ratio of this power to the power concentrated at the nominal and its harmonic frequencies. If this ratio is greater than some threshold, T, then the signal is classified as an audio signal. On the other hand, if this ratio is small meaning that the power is highly concentrated on the nominal frequency and its harmonics, then it is detected as a power signal.

B. Pole Estimation

Having identified the data type and its nominal frequency, next we estimate the pole locations of the data. Our dataset comprises of sinusoidal interference from the power grid, which is obscured by noise. As we want to estimate the sinusoidal frequency, intuition says an all pole model will work better. But since the data is non-stationary, it is not possible to model the system using the whole data. So, in order to fit such a model, the first challenge that we need to overcome is to select the segment length of the data. As we are interested in capturing the fluctuating frequency of the power grid, segmenting data becomes a vital issue. If we take a segment so long that the frequency fluctuates several times within it, our model will not give the accurate frequency estimation. Let the sampling frequency is f_s. So, if it is assumed that ENF remains more or less stationary for

t seconds, then $t * f_s = N$ samples form one block. So if the given signal is represented by $y(f)$ where $f = 0, 1, 2, ...$ denotes sample index, then the given data are partitioned into successive blocks as

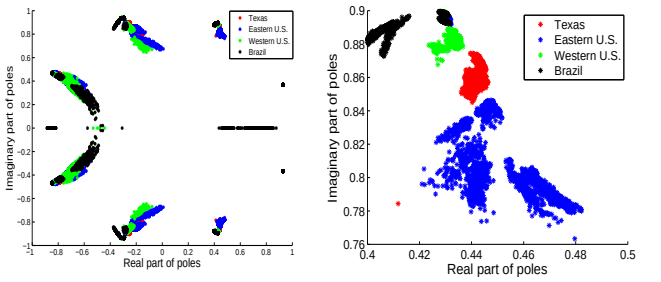
$$\begin{aligned} \mathbf{y}^q &= [y(Wq)\ y(Wq+1)\ ...\ y(Wq+N-1)]^T \\ &= [y^q(1)\ y^q(2)\ ...\ y^q(N)], \end{aligned} \tag{3}$$

where $q = 0, 1, 2, ...$ denotes block index and $\mathbf{y}^q$ represents the q^{th} block of the given data. Moreover, N and W indicate the block length and increment step between successive blocks, respectively. Next challenge is to select the model order m. As our data contains interference from the power line, it may have interference at harmonics of the nominal frequency due the non-linear nature of the interference. So, increasing the order of the model, information is obtained about the harmonics which is an important feature to identify a grid. Next we fit the data segment in the following model,
$y^q(n) = -a(2) * y^q(n-1) - a(3) * y^q(n-2) - \ldots - a(m+1) * y^q(n-m) + e(n)$,
where, $a(2), a(3)\ \ldots,\ a(m)$ are the AR model coefficients; $n = 1, 2, ...N$; m is the model order and e is model estimation error. Solving for the least square model estimation error, we get the following Yule-walker equation:

$$\begin{bmatrix} r(1) & r(2)^* & . & . & . & r(m)^* \\ r(2) & r(1) & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & r(2)^* \\ r(m) & r(2) & . & . & . & r(1) \end{bmatrix} * \begin{bmatrix} a(2) \\ a(3) \\ . \\ . \\ . \\ a(m+1) \end{bmatrix}$$

$$= \begin{bmatrix} -r(2) \\ -r(3) \\ . \\ . \\ . \\ -r(m+1) \end{bmatrix}$$

where, $\mathbf{r} = [r(1)\ r(2).....r(m+1)]$ is an autocorrelation estimate for y^q. Solving this we get the all pole model coefficients.



(a) Plot over the whole z-domain (b) Pole locations zoomed in to one cluster

Fig. 7. Scatter plot of the estimated poles for 60 Hz grids.

Figure 7a and 7b demonstrate the plot of the estimated poles for the power training data given for the 60 Hz grids. The zoomed in plot in Fig. 7b clearly shows the separate locations of the poles corresponding to different grids, which can be used to classify an unknown signal.

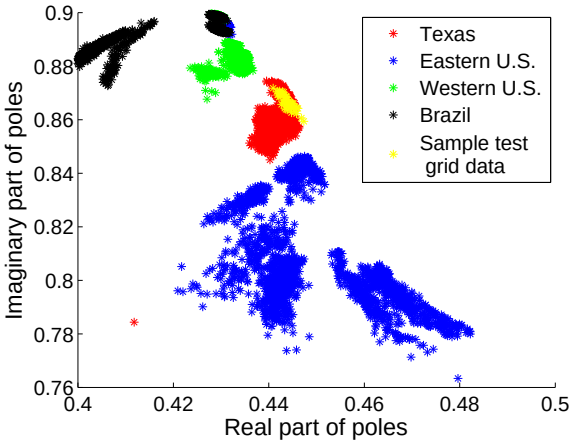


Fig. 8. Plot of the location of the poles of a testing set superimposed on the pole plot of the training grids.

C. Pole-matching Classifier

1) Training: In the training phase, poles are extracted from all the grids by the method mentioned above. These pole data are stored as ground truth to be compared with the testing data set in the testing phase.

2) Testing: In testing phase, first the audio data is extracted and segmented into different windows and then the pole data are estimated. After that, the main classification phase is initiated by measuring distances between the pole point of testing signal and the pole points stored in a database in training period previously. Let, p_i is the i^{th} pole of the testing data and $\mathbf{g}^k$ is the vector containing all the pole points of the training data of grid k. Then, a distance function is calculated as follows,

$$d^k_{i,j} = ||p_i - g^k_j||, \tag{4}$$

where $i = 1, 2, ...U$, $j = 1, 2, ...V$ and $k = 1, 2, ...R$. Again, U, V and R are total no. of testing poles, total no. of training poles and total no. of training grids, respectively. For each testing pole, the lowest X no. of distances are stored in a vector. Thus, a vector $\mathbf{d}^k$ of length $X * U$ is obtained for the grid k. We then calculate the average distance o^k from the distance vector as

$$o^k = \frac{1}{X * U} \sum_{l=1}^{X*U} d^k(l). \tag{5}$$

Then the final result for the test data is determined from the grid that gives the minimum distance. So, the estimated grid is,

$$\widehat{grid} = \arg\min_k (o^k). \tag{6}$$

An example is shown in Fig. 8. Here, the pole locations of a test signal is plotted along with the training poles. Clearly, the test poles (in yellow) match mostly with the poles of the Texas grid (marked as red). So, the grid location for this test set is Texas.

VI. Experimental Results

A. Parameter Settings

In the database used, we have total $R = 12$ number of grids. The sampling frequency f_s was 1000 Hz. We use $t = 10$ second windows for segmentation. As a result, block length is

TABLE II
CLASSIFICATION ACCURACY ON DIFFERENT GRIDS FOR POLE-MATCHING FOR THE TESTING SET. HERE THE DIGIT IN THE PARENTHESIS DENOTE THE TOTAL NO. OF FILES OF THAT TYPE AND THE DIGIT BEFORE THE PARENTHESIS DENOTE THE NO. OF CORRECTLY IDENTIFIED DATAFILE

Grid	No. of Power files	No. of Audio files	Accuracy (%)
A	9(9)	3(3)	100
B	7(7)	6(6)	100
C	7(8)	7(7)	93
D	6(7)	7(8)	87
E	7(8)	3(5)	77
F	8(8)	3(3)	100
G	9(9)	6(6)	100
H	8(8)	5(5)	100
I	7(8)	6(7)	87
J	4(4)	0(0)	100
K	4(4)	0(0)	100
L	4(4)	0(0)	100

$N = 10000$ and no overlapping is used in the algorithm, so $W = 10000$. Model order $m = 8$ was implemented for AR-parameter estimation, although the effect of this was analyzed in a later subsection by varying the order. Again the mean of the 2 closest training poles of each grid for each test pole was used for grid identification, so $X = 2$ was used.

B. Performance Measurement

The proposed algorithm's performance is evaluated using the accuracy criterion defined as

$$Accuracy(\%) = \frac{No.\ of\ correctly\ classified\ grids}{Total\ no.\ of\ grids\ classified} \quad (7)$$

We also consider the computational time for classification purpose in evaluating the algorithm.

C. Classification Accuracy

The primary focus of the region-of-recording identification task is to correctly locate the place where the media recording took place. So, the prime index of interest here is the classification accuracy. Among the 45 development data signals, the proposed algorithm is able to correctly identify 44 grids, which corresponds to 97.78% accuracy. On the 134 unknown test datasets, pole-matching is able to provide accurate estimation for 126 signals, resulting in an accuracy of 94.03%. Thus, the overall combined accuracy achieved for development and test set together is 94.97%. The detailed grid-wise classification results on the test set are given in Table II. From here, it can be observed that on most cases, the algorithm is able to correctly locate the grid. Interestingly, classification is extremely accurate for some of the grids such as grid A, B, H etc. compared to the other ones. This may be due to the fact that, on the z-domain, there is higher separability of the pole locations for those grids with others. Again, in some cases, poles of different grids overlap over some region due to the matching of pattern of their original grid signal which might hamper the accuracy. However, pole-matching is still able to maintain a satisfactory level of estimation success since the results are obtained through distance calculation of the training poles closest to the test poles. Thus, even if a large portion of training poles of 2 grids overlap, the algorithm provides quite accurate results as long as the test poles do not fall exactly in between the overlapping region of the training poles.

TABLE III
CLASSIFICATION ACCURACY ON DIFFERENT GRIDS FOR AN ENF-BASED SVM CLASSIFIER [14] FOR THE TESTING SET. HERE THE DIGIT IN THE PARENTHESIS DENOTE THE TOTAL NO. OF FILES OF THAT TYPE AND THE DIGIT BEFORE THE PARENTHESIS DENOTE THE NO. OF CORRECTLY IDENTIFIED DATAFILE

Grid	No. of Power files	No. of Audio files	Accuracy (%)
A	9(9)	2(3)	92
B	7(7)	6(6)	100
C	5(8)	3(7)	53
D	6(7)	8(8)	93
E	7(8)	2(5)	69
F	6(8)	3(3)	82
G	6(9)	2(6)	53
H	8(8)	4(5)	92
I	6(8)	7(7)	87
J	0(4)	0(0)	0
K	0(4)	0(0)	0
L	0(4)	0(0)	0

TABLE IV
CLASSIFICATION ACCURACY (%) ON DIFFERENT GRIDS FOR DIFFERENT ENF-BASED CLASSIFIERS [15] ALONG WITH POLE-MATCHING RESULTS

Classifier	P_50	P_60	A_50	A_60	Overall Accuracy
Pole-matching	**97.3**	94.74	**93.2**	**95.65**	**94.97**
KNN	79	92	62	88	78.31
RF	96	98	87	87	92.54
LP	90	97	60	78	82.08
NN	95	98	76	83	88.91
SVM	90	**99**	63	84	84.02

For comparison purpose, the dataset was also classified using the state-of-the art ENF-based SVM classifier, as described in [14]. The features were selected as given in [14] and the results are shown in Table III. The overall obtained accuracy for this method was 73.13%. To further compare the results obtained using pole-matching with those achieved using an ENF-based machine learning classifier, we compare our results with the reported results in [15]. Šarić *et al.* reported the results on the same database using five different ENF-based machine learning classifiers [15]. We denote the 4 types of signals, namely 50 and 60 Hz, power and audio as P_50, P_60, A_50 and A_60, respectively. The classification results are given in Table IV. The different classifiers used were K^{th} Nearest Neighbour (KNN), Random Forests (RF), Linear Perceptron (LP), Neural Network (NN) ans Support Vector Machines (SVM). Table IV depicts the performance of each of them on the 4 types of signals. Interestingly, the results reported are the best possible results obtained by tuning the parameters for all 4 types [15]. A consistent set of features were not used to obtain these results since in that case, accuracy dropped quite heavily [15]. Notably, the accuracy reported here for

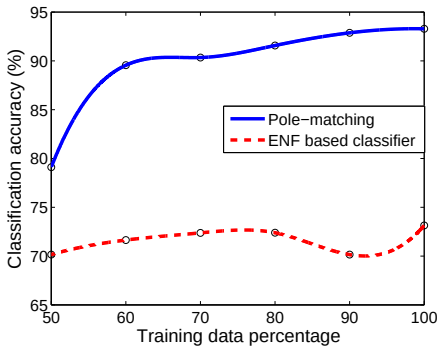


Fig. 9. Classification accuracy versus training data length for pole-matching and an ENF based SVM classifier.

the SVM classifier is higher than the one given in Table III since in [15], improved results were obtained for SVM using extra features. For each type, the best accuracy obtained is highlighted in Table IV. Notably, the proposed pole-matching classifier gives higher accuracy on most types and its overall accuracy is highest. In terms of overall accuracy, only the RF classifier gives comparable performance. However, it has been reported that its the slowest among all the ones given [15]. In a later subsection, we analyze that the proposed algorithm is computationally much faster than the SVM classifier. Since the RF is reported to be even much slower compared to SVM, clearly pole-matching provides a better option as a region-of-recording classifier compared to the state-of-the art reported classification results made based on ENF.

D. Robustness

Next, we check the consistency of the obtained results against changes in the parameters and analyze various effects.

1) Effect of Length of Training Data on Accuracy: Next we investigate the effect of training data length on classification performance. This holds significance since the volume of available data on the grids may vary and a robust algorithm is expected to work satisfactorily up to a certain limit even if the volume is lesser. So we changed the training data length from 50% to 100% and the results are shown in Fig. 9. As we can see, up to data length of 60% of the total available training data, the accuracy remains around or above 90%, but when the volume of available data becomes even lesser, accuracy starts to fall. However, the algorithm provides robust estimations over a wide range of training data volume. Again, this may be explained as- when the training data length is decreased, the number of training poles also decreases. However, since the final prediction is based on the distance between a test pole and its nearest two training poles averaged over all the test poles, even if some poles which are away from a test pole are not included, it does not effect the algorithm greatly, as long as the closer poles are there. However, with decrease in data, gradually there is lesser number of training poles available to match with and below a certain limit, accuracy starts to fall as the probability of disappearance of the actual closer training poles goes up. As a result, accurate comparison between the grids cannot be made and as such, error occurs. Again, the results obtained by varying the training data length using ENF-based SVM classifier is also shown in Fig. 9. We observe that for different data volumes, the proposed method consistently provides better estimations compared to the SVM classifier.

TABLE V
EFFECT OF LENGTH OF TESTING DATA ON THE ACCURACY OF POLE-MATCHING AND SVM.

Length of testing data (%)	Pole-matching Accuracy (%)	SVM Accuracy (%)
90	94.03	68.66
80	94.03	71.64
70	94.03	68.66
60	94.03	67.91
50	94.03	67.91
40	93.28	43.44
30	92.54	31.72
20	91.04	15.03
10	90.30	13.83

TABLE VI
EFFECT OF POLE ORDER CHANGE ON THE ACCURACY AND COMPUTATIONAL TIME.

Pole Order	Accuracy (%)	Mean computational time (s)
4	79.85	0.603
6	93.28	0.6318
7	92.54	0.6112
8	94.03	0.7354
9	93.28	0.7777
10	88.81	0.8031

2) Effect of Length of Testing Data on Accuracy: In our experimental data, testing data signals were of 10 min duration. However, in real-life scenario, the length of the signal available to track a certain location may vary and length may be quite small as well. So, to further explore the efficacy of the proposed method against variation in the length of testing data, we curtailed the testing signals in various proportions and determined the classification accuracy. The results are given in Table V. Interestingly, the results obtained using the pole-matching classifier is quite consistent even when the testing data are as small as just 1 min duration. For comparison purpose, results for the SVM classifier is also given. Notably, the pole-matching classifier is specially effective when the media clip to be tested is of very short duration. As we can see, when the testing data was taken of below 5 mins, result of SVM dropped sharply as for such a short duration, accurate feature extraction and classification becomes quite difficult. However, since even for a very short duration signal, we obtain some testing poles and the algorithm matches the location of those poles with the training ones, the pole-matching algorithm does not suffer from the testing data available for a limited time. This could be very useful for crime detection purposes where an audio clip from the criminals may be of just 1 or 2 mins and in those cases, the proposed pole-matching method is still able to identify the grid location.

3) Effect of Pole Order Change: The model order used to estimate the poles was varied and the results are shown in Table VI. It can be seen that the accuracy is quite consistent over the range of orders and the computational time is also lower for all cases compared to the time required to classify using the SVM classifier. The reason is that the estimation is made by calculating the two closest distance from each

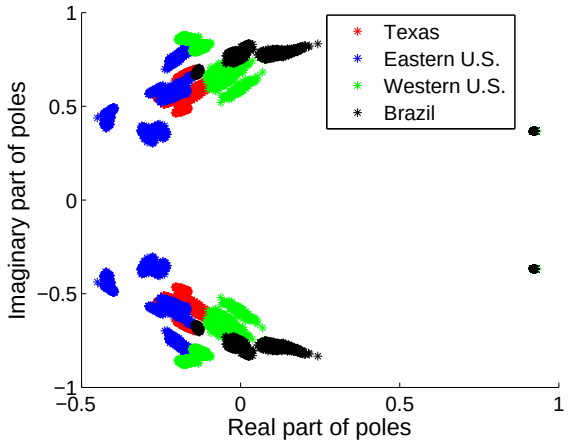


Fig. 10. Plot of the location of the poles for the 60 Hz training grids using model order 4 for AR parameter estimation.

test pole to the training poles. Thus if order is increased both for training and testing pole estimation, the number of pole clusters in the z-domain increases. Since matching is done with the nearest poles within the corresponding cluster, the mean distance does not change by a great extent. The pole plot with model order 4 is given in Fig. 10. The number of pole clusters has decreased compared to 7a as a lower pole order has been chosen.

E. Computational Complexity

Another important parameter of evaluation is the computational load. We used a MacBook with Intel core i5, 2.7 GHz processor and 8 GB RAM. The average time taken to classify a test recording of 10 min duration is 0.7354 ± 0.02 s (mean$\pm$ SD). However, the same measure for the ENF-based SVM classifier was 4.6055 ± 0.09 s. So, the proposed method is accurate as well as faster than the ENF-based method. Again, for the ENF-based SVM classifier, features have to be extracted and saved for the training data and as the number of feature increases, more memory is required but the pole-matching classifier works on a single pole location feature. Consequently, the computational complexity is much lesser.

F. Effect of Variation of ENF Window on SVM Classification Accuracy

To further investigate and compare the performances of pole-matching versus SVM accuracy against parameter changes, length of ENF window to extract the features for the SVM classifier was varied. In the pole-matching method, only time domain windowing is performed, whereas, for the SVM classifier, two-folds windowing is done. One is while extracting ENF from time domain signal and the other is windowing the extracted ENF signal to extract features from each ENF window. We experimented with ENF window size of 16, 32, 64, 96 and 128. The classification accuracy were 5.97%, 58.9552 %, 58.9552 %, 73.13% and 44.03%, respectively. So, the accuracy is not very consistent, rather depends on the choice of ENF window size. Again. the accuracy also varies with the choice of features such as level of wavelet decomposition. However, since parameters to be optimized for the proposed pole-matching approach are very few compared to the ENF-based classifier, the proposed method can provide more robust estimations.

G. Pole-matching Versus a Traditional ML classifier

For the classification purpose, here we have used a simple yet robust and effective distance based pole-matching classifier. However, if the training poles are used as features and a traditional SVM classifier is trained with these pole locations, then classification is also possible. But this gives rather poor performance with an accuracy of only 58.47%. Thus, the proposed classifier provides improved performance than if the poles are used to train a traditional machine learning (ML) classifier. Again, if a KNN classifier is used, then we get comparable performance with pole-matching. This is expected since both are based on distance matching. In case of the KNN classifier, the overall estimation accuracy is 90.30%. So the proposed method still provides slightly better performance.

VII. Conclusion

We have developed a robust HR estimation algorithm for heart rate monitoring during intensive physical exercises using wrist-type PPG signals. The motion artifact present in the PPG signal is successfully reduced by using a RLS-based multiple-reference adaptive denoising technique. Unlike conventional approach, the ANC generates four different versions of denoised PPG signal using the accelerometer data and the difference of the two PPG signals. Then the peak tracking stage intelligently selects the most suitable HR for the current window from the periodograms of these cleaned PPG signals. The experimental results show that our developed method can successfully estimate HR in real time even when the subject performs extensive physical exercise.

References

[1] *UNICEF: Child Protection From Violence, Exploitation and Abuse Child Trafficking.* [Online]. Available: http://www.unicef.org/ protection/57929_58005.html.

[2] R. Garg, A. Hajj-Ahmad, and M. Wu, "Geo-location estimation from Electrical Network Frequency signals," In *ICASSP,* pp. 2862–2866, 2013.

[3] N. Fechner and M. Kirchner, "Background Noise as a Carrier of ENF Artifacts in Mobile Device Audio recordings", 2014.

[4] C. Grigoras, "Digital audio recording analysis: The electric network frequency (ENF) criterion," *Int. J. Speech Lang. Law,* vol. 12, no. 1, pp. 63-–76, 2005.

[5] C. Grigoras, "Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis," *Forensic Sci. Int.,* vol. 167, pp. 136-–145, Apr. 2007.

[6] A. J. Cooper, "The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings an automated approach," In *33rd Audio Eng. Soc. Int. Conf. Audio Forensics-Theory and Practice,* Jun. 2008.

[7] A. J. Cooper, "An automated approach to the electric network fre- quency (ENF) criterion - theory and practice," *Int. J. Speech Lang. Law,* vol. 16, pp. 193-–218, Nov. 2009.

[8] C. Grigoras, "Applications of ENF analysis in forensic authentication of digital audio and video recordings," *J. Audio Eng. Soc.,* vol. 57, pp. 643-–661, Sep. 2009.

[9] A. J. Cooper, "Further considerations for the analysis of ENF data for forensic audio and video applications," *Int. J. Speech Lang. Law,* vol. 18, pp. 99-–120, Jun. 2011.

[10] R. Garg, A. L. Varna, and M. Wu, "Seeing ENF: Natural time stamp for digital video via optical sensing and signal processing," In *Proceedings of the 19th ACM international conference on Multimedia,* pp. 23–32, ACM 2011.

[11] Y. Liu, Z. Yuan, P. Markham, R. Conners, and Y. Liu, "Application of power system frequency for digital audio authentication," *IEEE Trans. Power Del.,* vol. 27, pp. 1820—1828, Oct. 2012.

[12] O. Ojowu, J. Karlsson, J. Li, and Y. Liu, "ENF extraction from digital recordings using adaptive techniques and frequency tracking," *IEEE Trans. Inf. Forensics Secur.,* vol. 7, pp. 1330-–1338, Aug. 2012.

[13] A. Hajj-Ahmad, R. Garg, and M. Wu, "Spectrum combining for ENF signal estimation," *IEEE Signal Process. Lett.,* vol. 20, pp. 885—888, Sep. 2013.

[14] A. Hajj-Ahmad, R. Garg, and M. Wu, "ENF-based region-of-recording identification for media signals," *IEEE Trans. Inf. Forensics Security,* vol. 10, no. 6, pp. 1125-–1136, Jun. 2015.

[15] Ž. Šarić, A. Žunić, T. Zrnić, M. Knežević, D. Despotović, and T. Delić, "Improving location of recording classification using Electric Network Frequency (ENF) analysis," In *Intelligent Systems and Informatics (SISY), 2016 IEEE 14th International Symposium on,* pp. 51–56. IEEE, 2016.

[16] Signal Processing Cup 2016, database. [Online]. Available: http://sigport.org/documents/information-mast-enf-power-signature-dataset.

[17] J. O. Smith, and X. Serra, "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," *International Computer Music Conference,* 2004.

[18] L. Dosiek, "Extracting electrical network frequency from digital recordings using frequency demodulation," *IEEE Signal Processing Letters,* vol. 22, no. 6, pp. 691–695, Jun. 2015.

[19] A. V. Oppenheim, and R.W. Schafer, *Discrete-Time Signal Processing,* Prentice-Hall, Englewood Cliffs, NJ, 1989, pp.713–718.