INSTANTANEOUS FREQUENCY ESTIMATION AND LOCALIZATION FOR ENF SIGNALS

Adi Hajj-Ahmad, Ravi Garg, and Min Wu

{adiha, ravig, minwu}@umd.edu
Department of Electrical & Computer Engineering
Institute for Advanced Computer Studies
University of Maryland, College Park, MD, USA.

ABSTRACT

Forensic analysis based on Electric Network Frequency (ENF) fluctuations is an emerging technology for authenticating multimedia recordings. This class of techniques requires extracting frequency fluctuations from multimedia recordings and comparing them with the ground truth frequencies, obtained from the power mains, at the corresponding time. Most current guidelines for frequency estimation from the ENF signal use non-parametric approaches. Such approaches have limited temporal-frequency resolution due to the tradeoffs of the time-frequency resolutions as well as computational power. To facilitate robust high-resolution matching, it is important to estimate instantaneous frequency using as few samples as possible. The use of subspace-based methods for high resolution frequency estimation is fairly new for ENF analysis. In this paper, a systematic study of several high resolution low-complexity frequency estimation algorithms is conducted, focusing on estimating the frequencies in short time-frames. After establishing the performance of several frequency estimation algorithms, a study towards using the ENF signal for estimating the location-of-recording is carried out. Experiments conducted on ENF data collected in several cities indicate the presence of location-specific signatures that can be exploited for future forensic applications.

Index Terms— Electric Network Frequency (ENF), Robust Instantaneous Frequency Estimation, Localization, Location-stamp

1. INTRODUCTION

Analysis of the Electric Network Frequency (ENF) signal is a recently proposed tool for multimedia forensics [1] [2]. ENF is the supply frequency in power distribution networks and its nominal value is 50 or 60 Hz depending on the geographic location. An important property of the ENF signal is that its value fluctuates around the nominal value; for example, the range of fluctuations is roughly on the order of 50-100 mHz in the United States. These fluctuations are due to variations in the load on the power grid and generally can be considered as random. This property of the ENF signal has enabled its emerging use in multimedia forensic analysis, particularly for timestamp authentication and forgery detection in multimedia recordings.

The ENF signal is embedded in such multimedia recordings as audio and video at the time when the recording is conducted. Audio recordings made near power source can capture the ENF variations from the electromagnetic interference from power lines [1]. Similarly, indoor video recordings made under fluorescent or incandescent lightings can capture the ENF signal due to near invisible flickering of lightings [2]. Frequency estimation techniques can then be applied to extract the ENF signals from media files for forensic studies. One such forensic application is the determination of the time-of-recording of a given file by comparing the ENF signal extracted from the media file with the ENF signal obtained directly from the power mains. In another application, the ENF signals from audio and video tracks of a given video can be compared to determine if the audio and visual data were captured at the same time [2].

Accurately and reliably estimating ENF signals is of utmost importance in its application for forensic purposes. This is because the validity of results depends on the accuracy of the frequency estimates obtained using a relatively short segment of media and sensor recordings. ENF signals in different recordings may be of varying quality. For example, ENF signals embedded in audio recordings are of low signal-to-noise ratio as compared with the signal recorded directly from the power mains. Moreover, the temporal resolution in timestamp estimation in media recordings using ENF signals is limited by the duration of frames used for instantaneous frequency estimation. Using a short frame may improve the resolution at a cost of decreased reliability in obtaining frequency estimates, if an appropriate frequency estimation is not used. Similarly, limited temporal resolution may also affect the capabilities of a forensic examiner analyzing a possible forgery on ENF signals for different counter forensic scenarios [3].

In the first part of this paper, we examine several techniques for estimating the instantaneous ENF frequencies in a raw measurement signal under different noise conditions. We also examine the effect of varying the frame size on the reliability of ENF signal estimation. The number of frames can become quite large for long media files, so we focus our study on examining methods without overly high computa-

tional complexity. We study the performance of different frequency estimation methods on synthetic and power data.

After experimentally evaluating the performance of different frequency estimation methods, we explore a novel application of ENF signal analysis in pinpointing the location in which a media or sensor recording was done. Most of the existing work has established that the ENF signal across an interconnected power grid is similar at a given time period. This is due to the load balancing mechanism employed in the power grids [4]. Variations may still be present in the frequency fluctuations at different locations due to local changes in the load and the finite propagation speed of the effects of this load change to other parts of the grid [5]. In this paper, we study these effects by conducting experiments on ENF data collected across several locations within the Eastern grid of the United States. This is the first study exploiting the ENF signal in location-stamping. It will be shown in our preliminary analysis that there exists differences among recordings taken in different locations within the same interconnected power grid. Such a property of the ENF signal may facilitate its use in a new range of forensic applications to pinpoint the location where a media file was recorded. The location specific traces in the ENF signal may be useful in an efficient monitoring of future smart-grids by assisting in fault localization [6].

The remaining of the paper is organized as follows. In Section 2, we study different frequency estimation techniques for ENF signal analysis. In Section 3, we experimentally evaluate the performance of our examined methods for ENF signal estimation under different signal quality conditions and different frame sizes. In Section 4, we study the characteristics of ENF signals obtained from different locations and propose a method to extract location specific signatures. Section 5 concludes the paper.

2. ENF EXTRACTION METHODS

For ENF extraction, a source signal containing the ENF signal is divided into overlapping frames of the same size. A frequency estimation algorithm is used on each frame to estimate the instantaneous frequency present in the frame. The result would be a "frequency vector", containing the frequency estimate for each subsequent frame. Retrieving frequency vectors from power signals and audio/video signals is an important first step in the ENF analysis. In this section, we review and discuss a few estimation techniques. Broadly speaking, they fall into one of two categories: the non-parametric approaches and the parametric subspace-based approaches.

2.1. Non-Parametric Approaches

Non-parametric approaches do not assume any explicit model for the data. Most of such approaches are based on the Fourier analysis of a signal. The main methods considered in this section are based on the spectrogram and the time recursive iterative adaptive approach (TR-IAA).

The spectrogram is a widely used tool for visualizing the Short-Time Fourier Transform (STFT). STFT is usually used for signals with a time-varying spectrum, such as the ENF signal. After the signal is segmented into overlapping frames, each frame undergoes Fourier analysis to determine the frequencies present. The spectrogram is then defined as the squared magnitude of the STFT. It is usually displayed as a two-dimensional intensity plot, with the two axes being time and frequency.

The ENF signal in a frame can be considered to be a sinusoid of a frequency close to the nominal value, embedded in noise. So, the power spectral density (PSD) of it, estimated by the spectrogram method, should ideally have a peak at the frequency of the sinusoidal signal. We consider two spectrogram-based approaches: (a) the maximum energy approach with quadratic interpolation, and (b) the weighted energy approach.

Maximum energy with quadratic interpolation: In this approach, we begin with finding the frequency that has the maximum spectral power component. Directly choosing this frequency as the frame's frequency estimate leads to much loss in accuracy, because the spectrogram is computed for a finite number of discrete frequencies and the actual frequency with the maximum energy may not be among them. To improve the accuracy of the estimate, the number of Fast Fourier Transform (FFT) points can be increased at a higher computational cost. An alternative approach is to use quadratic interpolation for fitting the PSD points corresponding to the range about the discrete frequency with the maximum energy [7]. This interpolation is outlined below based on the approach in [8].

As the computed PSD of each frame is a function of the discrete frequencies, we can think of it as a function of frequency bin numbers, or simply indices. Denoting the index of the frequency with the maximum energy as k_{max}, we can define a coordinate system centered at $(k_{max}, 0)$. We take the log magnitude value of the PSD as $y(k)$. Using three points of the PSD, namely at $k_{max}-1$, k_{max} and $k_{max}+1$, we can carry out quadratic interpolation on the parabola of the form: $y(k) = a(k-p)^2 + b$. Solving for parabola peak p, we get the following expression.

$$p = \frac{1}{2}\frac{y(k_{max}-1) - y(k_{max}+1)}{y(k_{max}-1) - 2y(k_{max}) + y(k_{max}+1)}, \qquad (1)$$

where $y(k)$ is defined as

$$y(k) = 20\log_{10}|PSD(k)|. \qquad (2)$$

The estimate of the peak location in bins corresponding to the true frequency is $k_t = k_{max} + p$. The frequency estimate corresponding to k_t is $\frac{k_t f_s}{N}$, where f_s is the sampling frequency and N the number of FFT points used in computing the spectrogram.

A drawback of this approach is that it is susceptible to outliers. If the maximum energy happens to be far away from the nominal frequency due to additive noise or interference from content, the subsequent estimation would be erroneous.

Weighted energy: The weighted energy approach makes use of the additional information that we have, namely that the frequency should be close to the nominal frequency. As a result, the frequency estimates are robust to outliers in frequency [2].

The weighted energy approach finds the frequency estimate $F(n)$, for the n^{th} frame, by weighing the frequency bins around the nominal values. The expression for frequency estimates is then given by the following equation [2]:

$$F(n) = \frac{\sum_{l=L_1}^{L_2} f(n,l)S(n,l)}{\sum_{l=L_1}^{L2} S(n,l)} \tag{3}$$

where L_1 and L_2 are the FFT indices of the boundary of the averaging region and $f(n,l)$ and $S(n,l)$ are the frequency and energy in the l^{th} frequency bin of the n^{th} time frame.

TR-IAA: A class of non-parametric frequency estimation methods most recently developed is based on the time recursive iterative adaptive approach (TR-IAA) [7]. This algorithm reaches the spectral estimates of a given time frame by minimizing a quadratic cost function using a weighted least squares formulation. TR-IAA is an iterative technique, which takes from 10 to 15 iterations to converge, where the spectral estimate is initialized to be either that of the spectrogram or the final spectral estimate of the preceding time frame [7] [9]. This method is more computationally extensive than spectrogram-based techniques. After the convergence of the spectral estimate, a quadratic interpolation scheme similar to the first spectrogram-based approach discussed is used to estimate the frequency.

It has been shown that the TR-IAA based approach provides slightly better frequency estimates than the spectrogram based approach, when the frame size is 20-30 seconds [7]. Since we are considering short frame sizes and due to the high computational costs of the TR-IAA based algorithm observed, we focus on spectrogram-based approaches for non-parametric methods in this paper.

2.2. Parametric Approaches

Parametric methods assume an explicit model for the signal and underlying noise. Due to such an explicit assumption about the model, the estimates obtained using parametric approaches are more accurate than non-parametric approaches [10]. In this paper, we consider two of the most widely used parametric frequency estimation methods based on the subspace analysis of a signal and noise model. In general, these methods can be used to estimate the frequency of a signal composed of P complex frequency sinusoids embedded in white noise. As ENF signals consist of only one real sinusoid, the value of P for ENF signals is 2. The methods we study are the MUltiple SIgnal Classification (MUSIC) and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT). Next, we briefly describe these two methods [10].

MUSIC: The MUSIC algorithm is a subspace-based approach to frequency estimation that relies on eigen-decomposition and the properties between the signal and noise subspaces for sinusoidal signals with additive white noise. The algorithm first computes an $M \times M$ correlation matrix, where M is chosen to be larger than P, the number of anticipated complex exponentials. More specifically, for an N-point observed signal $x[n]$ where $N >> M$, we generate an $M \times N$ data matrix of the form:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}(1) & \mathbf{x}(2) & ... & \mathbf{x}(N-2) & \mathbf{x}(N-1) \end{bmatrix}^T \tag{4}$$

where $\mathbf{x}(n) = \begin{bmatrix} x(n) & x(n+1) & ... & x(n+M-1) \end{bmatrix}^T$.

An estimate of the correlation matrix $\hat{\mathbf{R}}_x$ can be computed by $\hat{\mathbf{R}}_x = \frac{1}{N}\mathbf{X}^H\mathbf{X}$, where the superscript H denotes the Hermitian of a matrix. Eigenanalysis is carried on $\hat{\mathbf{R}}_x$ to find the vectors spanning the signal and noise subspaces. These subspaces are orthogonal to each other. The eigenvectors $\mathbf{q}_1, \mathbf{q}_2, ..\mathbf{q}_P$ corresponding to the largest P eigenvalues that span the signal subspace and the remaining eigenvectors $(\mathbf{q}_{P+1}, \mathbf{q}_{P+2}, ..\mathbf{q}_M)$ span the noise subspace.

MUSIC makes use of the orthogonality property of noise eigenvectors and steering vectors $\mathbf{v}(f_k)$, $1 \leq k \leq P$ (corresponding to the actual frequency components of the signal). Here $\mathbf{v}(f)$ is defined as:

$$v(f) = \begin{bmatrix} 1 & e^{j2\pi f} & e^{j4\pi f} & ... & e^{j2(M-1)\pi f} \end{bmatrix}^T \tag{5}$$

The pseudo-spectrum of MUSIC, $\bar{R}_{music}$, can then be computed as:

$$\bar{R}_{music}(e^{j2\pi f}) = \frac{1}{\sum_{m=p+1}^{M} |\mathbf{v}^H(f)\mathbf{q}_m|^2} = \frac{1}{|Q_m(e^{j2\pi f})|^2} \tag{6}$$

Due to the orthogonality property, the denominator should be zero at the frequencies of the signal.

Among various techniques studied in the statistical signal processing literature, RootMUSIC provides high precision at a moderate computational cost. It solves for the roots of the denominator directly [10]. The frequency estimates using this method are the arguments of the P roots closest to the unit circle. We opted to use RootMUSIC in our comparisons.

ESPRIT: ESPRIT makes use of the rotational property between staggered subspaces that is invoked to produce the frequency estimates. In our case, this property relies on observations of the signal over two intervals of the same length staggered in time. ESPRIT is similar to MUSIC in the sense that

they are both subspace-based approaches, but it is different in that it works with the signal subspace rather than the noise subspace. The implementation that we employ estimates the signal subspace from the data matrix $\mathbf{X}$ of Eq. 4.

A Singular Value Decomposition (SVD) is applied to $\mathbf{X}$, giving:

$$\mathbf{X} = \mathbf{LSU}^H \tag{7}$$

where $\mathbf{L}$ is an $N \times N$ matrix of the left singular vectors, $\mathbf{S}$ is an $N \times M$ matrix with its main diagonal entries containing the singular values, and $\mathbf{U}$ is an $M \times M$ matrix of the right singular vectors. The singular values correspond to the square roots of the eigenvalues of the sample correlation matrix $\hat{\mathbf{R}}_x$ scaled by N, and the columns of $\mathbf{U}$ are the eigenvectors of $\hat{\mathbf{R}}_x$. These vectors form an orthonormal basis for the underlying M-dimensional vector space. More specifically, $\mathbf{U}$ can be written as $\mathbf{U} = [\mathbf{U}_s|\mathbf{U}_n]$, where $\mathbf{U}_s$ is the $M \times P$ matrix of right singular vectors corresponding to the singular values with the P largest magnitudes and $\mathbf{U}_n$ is the $M \times (M - P)$ matrix containing the remaining right singular vector. The signal subspace can be partitioned into two smaller $(M - 1)$-dimensional subspaces as:

$$\mathbf{U}_s = \begin{bmatrix} \mathbf{U}_1 \\ * \end{bmatrix} = \begin{bmatrix} * \\ \mathbf{U}_2, \end{bmatrix} \tag{8}$$

where $\mathbf{U}_1$ and $\mathbf{U}_2$ correspond to the unstaggered and staggered subspaces, respectively. The relation between $\mathbf{U}_1$ and $\mathbf{U}_2$ can written as:

$$\mathbf{U}_2 = \mathbf{U}_1\mathbf{Q}, \tag{9}$$

where $\mathbf{Q}$ is a $P \times P$ matrix. $\mathbf{Q}$ can be computed using a least squares method. Eigenanalysis can then be carried out on $\mathbf{Q}$. The frequency estimates can be extracted from the arguments of the eigenvalues of $\mathbf{Q}$. Denoting these eigenvalues by $\phi_k, 1 \leq k \leq P$, we can find the frequency estimates, $\hat{f}_k$, by:

$$\hat{f}_k = \frac{\angle\phi_k}{2\pi} \quad \text{with} \quad 1 \leq k \leq P. \tag{10}$$

It is worth noting that the accuracy of the estimates using subspace methods can differ significantly depending on the parameter M chosen for the data matrix in Equation 4. It was shown in [11] that for estimating the frequency in a single sinusoid in white noise, the error variance is minimal when M is either $\frac{N}{3}$ or $\frac{2N}{3}$. For a general multiple signal case, a rule of thumb suggested in [12] is that M should be in the range $[\frac{2N}{5}, \frac{3N}{5}]$. So, before we carry out the experiments detailed in the next section, we test the accuracy of MUSIC and ESPRIT on the range $M \in [\frac{N}{3}, \frac{2N}{3}]$ for the N that we intend to use. We find the value of M that gives the most accurate result for each case of N, and subsequently use it in the following experiments. The metric used for determining accuracy is the correlation coefficient, which will be discussed in the next section.

3. COMPARISON OF FREQUENCY ESTIMATION METHODS

In this section, we compare the performance of different frequency estimation methods discussed in Section 2. We present the experiments that we carry out and the criteria by which we compare the methods.

3.1. Experiments on Synthetic Signals

To facilitate a comparative study, we first generate synthetic signals with ground truth frequencies. In our model of the synthetic signal, the frequency of the signal changes on a frame basis and the frequency estimation algorithms are applied to each time-frame. Since the ENF signal has a slowly-varying frequency, we attempt to capture the correlation of the frequencies in consecutive frames by first generating a random sequence of frequencies having a normal distribution of mean $\mu = 60$ and standard deviation $\sigma = 0.0133$. Then, to mimic the pseudo-periodic behavior of the ENF fluctuations, we pass this sequence through a filter of the form [13]:

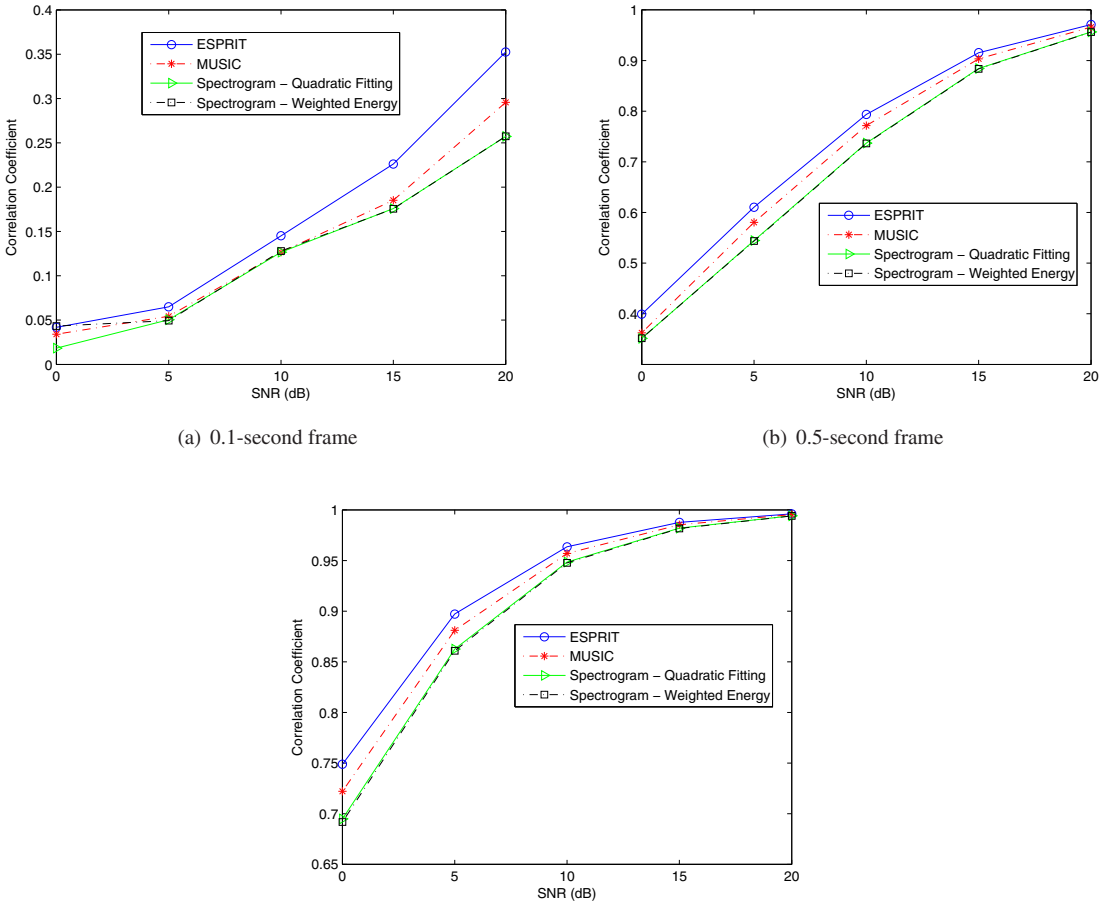
$$H(z) = \frac{1}{1 - 0.97z^{-1}}. \tag{11}$$

For one instance of this simulation, the synthetic signal is generated as a series of consecutive sinusoidal signals of frequencies corresponding to the generated sequence, each with a phase following a uniform distribution $U(0, 2\pi)$. We consider the sampling frequency of 441Hz, and examine several frame sizes, ranging from 0.1 seconds (44 samples) to 1 second (441 samples). We add additive white gaussian noise (AWGN) to achieve different levels of signal-to-noise ratio (SNR).

We carried out 100 simulation runs where each run has a different set of frequencies, phase angles, and additive noise. In each simulation run, the frequencies are estimated on a frame by frame basis using the frequency extraction methods discussed in Section 2. For the spectrogram-based approaches, the number of FFT points is chosen such that the frequency resolution is approximately 0.03 Hz. For the weighted energy approach, the range considered for weighting is [59.8, 60.2] Hz. For the subspace methods, we use the experimentally optimized M value for the dimension of the data matrix based on the discussions in Section 2.2.

A direct way of comparing the frequency estimates is to subtract each sequence of estimates from the true sequence of frequencies and estimate the mean difference. This criterion does not fit our application because it may penalize some methods that tend to have an inherent bias in estimating the frequency, such as the weighted energy approach. Since we are interested in measuring the similarity in the trends of the signals for most applications based on ENF signal analysis, these trends are better revealed using a correlation based metric.

More specifically, to measure the performance of different frequency estimation methods, we use the cross-correlation



(a) 0.1-second frame

(b) 0.5-second frame

(c) 1-second frame

Fig. 1. Synthetic Signals: Correlation Coefficient vs. SNR for very short frame sizes

between the frequency estimates and the true frequencies and measure the correlation for different frame sizes and different SNR values. The results are shown in Fig. 1. The correlation coefficient of two sequences that do not match is close to zero, and the correlation coefficient for matching sequences is expected to be higher. From Fig. 1, we can see that the frequency estimates worsen when more noise is added and the SNR decreases, which is expected. Also, when the frame size becomes small, as with Fig. 1(a), all estimation techniques behave rather poorly, as the correlation coefficient achieved was around 0.35 in high SNR cases and only 0.05 in low SNR cases; the latter is barely differentiable from matching two unrelated ENF signals. For this reason, it is advisable not to use such small frame sizes whenever possible.

We also note that both spectrogram-based approaches give similar performances; the subspace-based approaches give better performances than the spectrogram-based approaches, with ESPRIT consistently outperforming MUSIC by a moderate margin. This is due to an explicit assumption on the sinusoidal signal model in the parametric approaches and this signal model matches the synthetic model well. As with ESPRIT outperforming MUSIC, our observation in the context of the ENF problem is consistent with the general result in the literature that ESPRIT gives slightly more accurate estimates than MUSIC for a similar computational cost [14].

3.2. Experiments on Power Grid Data Set

To validate the results that we have obtained using synthetic data, we examine the performance on a power signal measurement available at [15]. This data set also provides a sequence of reference frequencies computed using a Frequency Disturbance Recorder (FDR). FDR estimates the frequencies using phasor analysis and signal resampling techniques, and the FDR measurement is known to provide a resolution of around ± 0.0005 Hz [16].

We assume the power measurement to be noise-less for simplicity, and we add various levels of AWGN to it. We estimate the frequencies present for different frame sizes using the estimation techniques discussed in Section 2. We use the

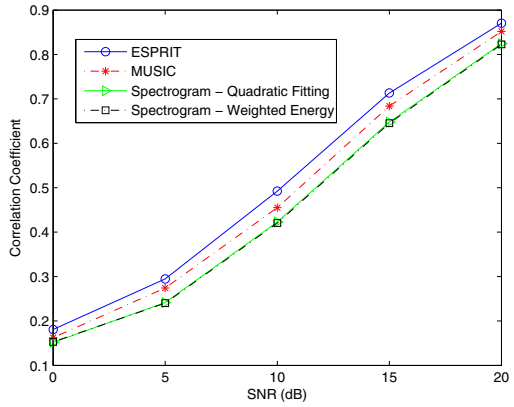


Fig. 2. Power Grid Data Set: Correlation Coefficient vs. SNR for frame size of 1 second

frequencies computed by the FDR as a reference, and compute the correlation coefficient between this reference and the estimated sequence of frequencies. This correlation was computed after temporally aligning the reference frequency sequence with the estimated ones.

The results for 1-second frame size can be seen in Fig. 2. The estimates that we obtained are worse than the estimates for synthetic data, which is understandable as the model we used was an idealization of the reality and the power measurements is likely not noise free or a perfect sinusoid. However, the order of accuracy remains the same as with the synthetic data comparison, with ESPRIT outperforming the other three methods, followed by MUSIC. Fig. 2 shows similar trend and relative positions to those shown in Fig. 1(c) except that the correlation coefficient values are lower.

3.3. Matching ENF Signals from Audio and Power

A main forensic application involving ENF analysis requires matching an ENF signal extracted from an audio signal to that extracted from a power signal in order to determine the time of recording. Here, we carry out experiments to compare which of the four estimation methods gives the best matching between ENF signals that are known to be recorded at the same time.

More specifically, we make two simultaneous recordings of audio and power signals. The power signal is obtained from an electric outlet using a step-down transformer. The audio signal is obtained by recording the background noises in a room. As mentioned in the introduction section, the audio recording may pick up the ENF signal from electromagnetic interferences in the room. We expect that if there is ENF present in the audio, it should match with that in the power recording captured concurrently.

Prior to ENF extraction, we pass the audio signal through a bandpass filter centered around 60Hz, the nominal frequency, to remove as much noise as possible without affecting the frequency band of interest. We partition both the power signal and the filtered audio signals into overlapping frames, and compute the frequency estimates for each frame using the four ENF extraction techniques under study. In order to differentiate the various estimation methods, we compute the correlation coefficient between the audio and power estimates of each method.

Table 1. Correlation coefficient values between the ENF signals extracted from the power and audio recordings.

Method	Correlation Coefficient
ESPRIT	0.9778
MUSIC	0.9767
Spectro. – Quad. Interp.	0.9704
Spectro. – Weighted Energy	0.9543

The results are shown in Table 1, where a higher value of the correlation coefficient suggests better matching between the ENFs extracted from the audio and power signals. Although the values in Table 1 are close, they support the findings reached earlier that ESPRIT gives the best results followed by MUSIC. This table also demonstrates to us that the spectrogram-based quadratic interpolation approach performs better than the weighted energy approach. This can be due to our restricting the estimates of the weighted energy to a chosen range which may not be true in all signal cases. The quadratic interpolation approach has no such restraints.

4. ENF SIGNAL AS A LOCATION STAMP

We have demonstrated in Section 3 the performance of different frequency estimation algorithms for ENF analysis using a correlation based metric. In this section, we explore a novel application of ENF signal analysis for spatial localization of sensing signals. The ENF signal has been used to determine the time of recording of a given multimedia signal that contains traces of ENF variations during the sensing process [1] [2]. Coarse location information in terms of the grid in which the recording was made may also be determined using ENF analysis. The specific ENF variations are not the same for recordings made over different power grids that are not frequency synchronized, for example, in the east interconnection grid versus the west interconnection grid. Different degrees of controls in regulating the power grid also play a key role [4]. The range of fluctuations in the ENF signal also generally depends on the size of the grid, as a smaller capacity grid often exhibits a higher range of fluctuations, while a large capacity grid can be more tightly controlled and exhibits a smaller range of fluctuations. For example, ENF variations in the UK power grid are over a higher range than the conti-

nental European Interconnection grid [4].

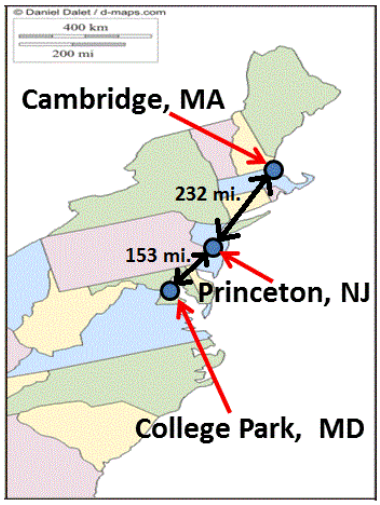


Fig. 3. Map of the locations where ENF recordings were conducted. [Source of base map: www.d-maps.com]

In addition to inter-grid variations, ENF signals may also exhibit intra-grid variations. These variations among ENF signals within the same grid are due to the local load characteristics of a given city and the time taken to propagate the response to demand and supply in load to other parts of the grid [4]. In general, power demand and supply in a given area follows a cyclic pattern; for example, demand increases during evening hours in a residential neighborhood. Any change in the load in a grid is regulated by a control mechanism [4]. An increase in the load causes the supply frequency to drop temporarily; the control mechanism senses the frequency drop and starts drawing power from adjoining areas to compensate for the increased demand. As a result, the load in adjoining areas also increases, which leads to a drop in the instantaneous supply frequency, and the overall power supply will be driven up to compensate the rising load which leads to a drop in the instantaneous supply frequency in those regions. A similar mechanism is used to compensate for an excess supply of power flow that leads to a surge in supply frequency.

A small change in the load in a given area may have a localized effect on ENF in the given area. However, a large change, such as one caused by a generator failure may have an effect on the whole grid. These changes are known to propagate along the grid at a typical speed of approximately 500 miles per seconds [5]. We conjecture that small and large changes in the load may cause location specific signatures in the ENF patterns, and if so, such differences may be exploited to narrow down the place of a given recording within a power grid. Due to a finite speed of propagation of frequency disturbances across the grid, we anticipate that the ENF signal would be more similar for locations close to each other as compared with the locations farther apart. Such a property

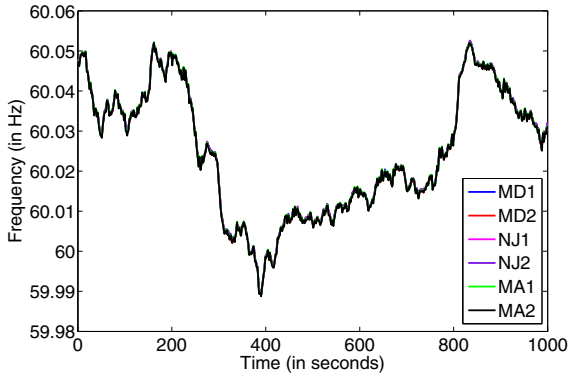


Fig. 4. Aligned ENF signals recorded in MD, NJ and MA (best viewed in color)

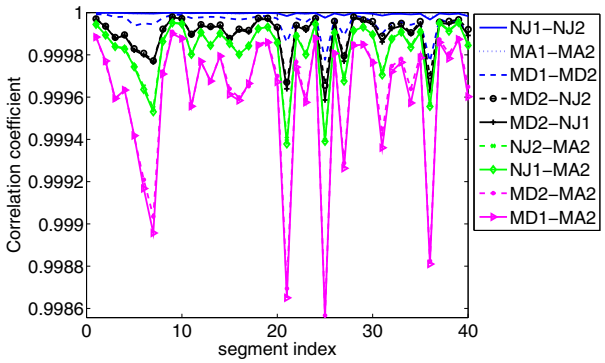
of the ENF signal propagation across the grid can be potentially used for localization at a finer resolution within a grid by comparing the similarity of the ENF signal in question with the ENF databases that may be available for a set of locations.

Devising a methodology to pinpoint the place of recording is useful for forensic analysis on the media and sensor recordings although locating within a grid using ENF traces has not been demonstrated in the literature to the best of our knowledge. In this section, we explore the presence of location specific signatures based on the experiments on ENF signals obtained from three locations in the Eastern interconnection grid of the United States.

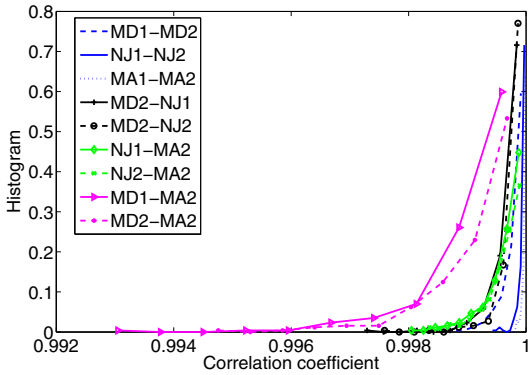
4.1. Experiments Conducted and Observations

We record power signals in Maryland (MD), New Jersey (NJ), and Massachusetts (MA) in the United States. These states are part of the Eastern interconnection grid. As shown in the map of Fig. 3, the recordings are conducted in College Park in MD, Princeton in NJ, and Cambridge in MA. The approximate distance between College Park and Princeton is 153 miles, while the distance between Cambridge and Princeton is 232 miles. Power signals are recorded by providing a step-down voltage from power-mains supply as an input to the sound card of Olympus WS700 series audio recorders. The sampling frequency of the Olympus recorders is 44.1 kHz, and it does not provide alternative sampling rates. At each location, we conduct two separate recordings at the same time at close vicinity to each other. We record power signals from these locations for multiple days. All recordings are started simultaneously by synchronizing with a webclock [17].

Each recorded power signal is first lowpass filtered and downsampled to 250 Hz. A bandpass filter with passband in 59-61 Hz is then used to extract the ENF signal at 60 Hz. Instantaneous frequency is estimated using the ESPRIT method by dividing the signal into 2-seconds long frames, and each frame has a 1-second overlap with the previous frame. Al-



(a) Correlation coefficient



(b) Histogram of Correlation coefficient

Fig. 5. Correlation coefficient between ENF signals obtained across different locations for 500-second query segment length.

though the recordings were manually synchronized using a web clock at [17], misalignment of the ENF sequences may still be present. As the main trends of ENF signals are known to be the same at locations within a grid, we temporarily align all ENF sequences by finding the time corresponding to the maximum value of their pairwise correlation coefficient. Fig. 4 shows a plot of a 1000 samples of aligned ENF sequences recorded on 20-21 July 2012 across the three locations described above. From this figure, we observe that as expected, the ENF signals at a given time are similar at different locations across the same grid. We compute the pairwise correlation coefficient between the recorded ENF signals for segments of 500-samples long. Each sample is the frequency estimate of a 2-second frame, with 1-second overlap with the previous frame, as explained earlier. We plot these correlation coefficients in Fig. 5(a) and its histogram plot is shown in Fig. 5(b). From these figures, we observe that the ENF sequences across different cities in the same grid recorded at the same time are highly correlated with each other.

It is difficult to distinguish the signals recorded at different locations using such a macroscopic correlation plot. However, a careful "microscopic" examination reveals some differences in the ENF sequences. A zoomed plot of an ENF sequence from a recording from each of the three locations is shown in Fig. 6. From this figure, we observe that the ENF sequences at each location have different location specific traits, which are not present at other locations. Location-specific traits are present in the recordings conducted across all three cities. Next, we explore such location-specific signatures.

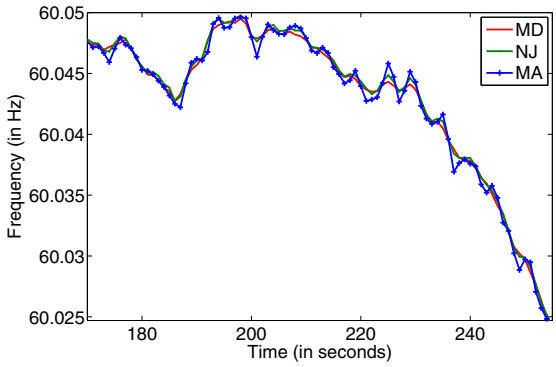


Fig. 6. Zoomed view of the ENF signals in different states (best viewed in color)

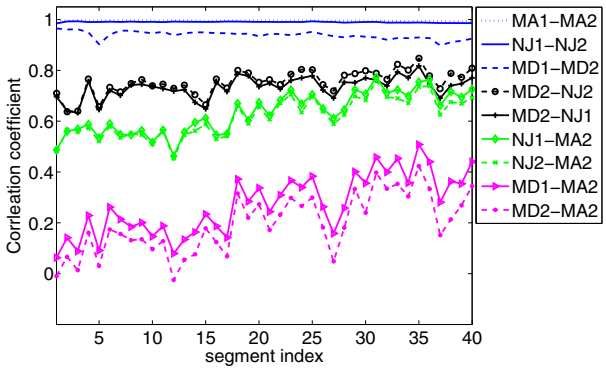
4.2. Exploring Location Signatures

In this subsection, we describe a method used to extract the location-specific characteristics from each recording. As shown in Fig. 6 earlier, the microscopic trends in the ENF signal recorded at distinct locations are different. These detailed trends vary fast with time, and are present at higher frequency details in the ENF signal. This motivates us to apply a high-pass filter to extract such location specific signatures. To realize such a high-pass filter, we pass each ENF signal through a smoothening filter, and subtract the resulting signal from the original ENF sequence. The highpass filtered ENF signal, $f_{hp}(n)$, can be written as:

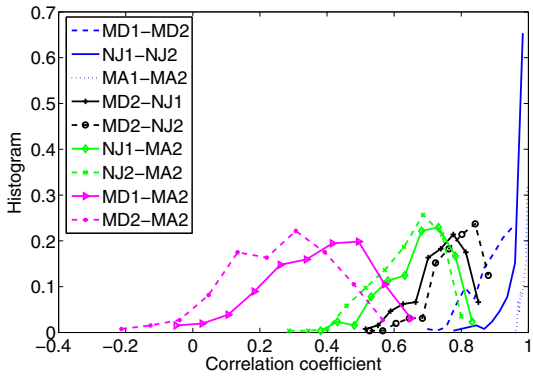
$$f_{hp}(n) = f(n) - \sum_{k=-\frac{N-1}{2}}^{\frac{N-1}{2}} w(k) f(n-k) \quad (12)$$

where $f(n)$ is the ENF value at time n, $w(\cdot)$ is the coefficient of the smoothening filter, and N is the order of the filter. We choose an odd order for the filter.

After obtaining the high-pass filtered ENF sequence for each recording, we use the correlation coefficient of the high-pass filtered ENF sequences as a metric to examine the location similarity between two recordings. If the proposed filter is able to capture the location specific signatures, we expect to obtain a high value of correlation coefficient for intra-city recordings, and a low-value for inter-city recordings.



(a) Correlation coefficient



(b) Histogram of correlation coefficient

Fig. 7. Correlation coefficient between ENF signals across different locations for 500-second query segment length when high-pass filtered ENF signal is used for location matching.

4.3. Results and Discussions

We now discuss the results obtained from a 14-hour long recording of the power ENF signal conducted across the three different locations described above, namely, College Park, Princeton, and Cambridge/Boston. After aligning the estimated ENF sequences temporally, we pass each signal through a smoothening filter of order $N = 3$. We use a uniform weight vector $w(k) = \frac{1}{N}$ for all k. This is followed by subtracting the smoothed version from the original ENF signal to obtain a highpass filtered ENF signal. We divide the high-pass filtered signals into non-overlapping segments of 500 samples each. The correlation coefficient of each segment for the recordings across different cities is shown in Fig. 7(a). From this figure, we observe relatively higher values of correlation coefficient for all intra-city recordings (MD1-MD2, NJ1-NJ2, MA1-MA2) for the top three curves as compared with inter-city recordings.

We further plot the histogram of correlation coefficients between different highpass filtered ENF signals across different locations in Fig. 7(b). From this figure, we observe that the pairwise correlation coefficient between intra-city recordings is close to 1, and for inter-city recordings, it is much less than 1. This observation suggests that the highpass filter can potentially capture the city specific signatures from the ENF signal.

Among the inter-city data, we observe that the value of correlation coefficient for MD-NJ and NJ-MA recordings are higher than those for MD-MA recordings. As can be seen earlier in Fig. 3, Princeton, NJ, is located approximately midway between College Park, MD, and Cambridge, MA, and so its distance to College Park and Cambridge is smaller than the distance between College Park and Cambridge. A high value of correlation coefficient for MD-NJ and NJ-MA recordings as compared to the correlation coefficient between MD-MA recordings reflects the time delay from the propagation effect of the load balancing, as was explained in Section 4. We observe from Fig. 7(a) that the values of correlation coefficient for MD-NJ recordings are moderately higher than NJ-MA recordings. This reflects the smaller distance between College Park and Princeton as compared to the distance between Princeton and Cambridge as shown in Fig. 3. These plots demonstrate that the ENF signal has a strong potential to be used as a location stamp by comparing the correlation coefficient of the highpass filtered query ENF signal with the highpass filtered ENF signal database of each city from the corresponding time.

From the three-city data for the U.S. east coast, we observe that the correlation coefficient between the highpass filtered query ENF signals captured across different cities in a geographical area covered by a grid is approximately proportional to the distance between the locations. As a result, the correlation coefficient can be used as an indicator of the distance between two cities. Given the distance estimated between cities and the known location-of-recording of some ENF signals behaving as anchor signals, trilateration techniques may be used to estimate the recording location of a query. More generally, however, it may become difficult to characterize a simple relationship between correlation coefficients and distances, as the correlation coefficient may depend on such factors as power line density in a grid, and wired distances may vary from the flight or road distances between the cities. To mitigate this issue, we are collecting data from multiple locations in the eastern and the western grid of the United States to empirically characterize correlation coefficient profiles of the grid with geographical distance. Such a study will help in a better understanding of the resolution limits of the location stamping properties of the ENF signals.

5. CONCLUSION AND FUTURE WORK

In this paper, we have studied different parametric and non-parametric frequency estimation methods with application to Electric Network Frequency (ENF) signal estimation in high temporal and frequency resolution. We have conducted ex-

periments on synthetic data and experimental data under different noise conditions to evaluate the performance of studied methods using a correlation coefficient based metric. Our results have demonstrated that the ESPRIT based parametric frequency estimation method provides the best results for ENF matching using correlation coefficient. We have also studied a novel application of ENF signal analysis in pinpointing locations based on the characteristics of ENF signals recorded in several cities of the United States Eastern interconnection power grid. We have proposed a method based on high-pass filtering to extract location traces in the ENF signal. Preliminary results of the proposed approach demonstrate that the ENF signal carries location specific traces that have a strong potential towards estimating the location of media and sensor recordings.

Our ongoing and future work includes conducting a more comprehensive study of ENF signal based location-stamping by collecting more ENF data across different power-grids in different cities and different sensing modalities.

Acknowledgment

The authors would like to thank Dr. Shan He, Mr. Chau-Wai Wong, and Mr. Michael Luo for their help in the power data collection for the research work done in this paper.

6. REFERENCES

[1] C. Grigoras, "Applications of ENF criterion in forensics: Audio, video, computer and telecommunication analysis," *Forensic Science International*, vol. 167, no. 2-3, pp. 136 – 145, Apr. 2007.

[2] R. Garg, A. L. Varna, and M. Wu, "'Seeing' ENF: natural time stamp for digital video via optical sensing and signal processing," in *Proeedings. of the 19th ACM International Conference on Multimedia*, 2011, pp. 23–32.

[3] R. Garg W. H. Chuang and M. Wu, "How secure are power network signature based time stamps?," *ACM Conf. on Computer and Communication Security*, Oct. 2012.

[4] M. Bollen and I. Gu, *Signal Processing of Power Quality Disturbances*, Wiley-IEEE Press, 2006.

[5] S. J. Tsai, L. Zhang, A.G. Phadke, Y. Liu, M.R. Ingram, S.C. Bell, I.S. Grant, D.T. Bradshaw, D. Lubkeman, and L. Tang, "Frequency sensitivity and electromechanical propagation simulation study in large power systems," *IEEE Transactions on Circuits and Systems*, vol. 54, no. 8, pp. 1819 –1828, Aug. 2007.

[6] S. Massoud Amin and B.F. Wollenberg, "Toward a smart grid: power delivery for the 21st century," *IEEE Power and Energy Magazine*, vol. 3, no. 5, Sep. 2005.

[7] O. Ojowu, J. Karlsson, J. Li, and Y. Liu, "ENF extraction from digital recordings using adaptive techniques and frequency tracking," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, pp. 1330 – 1338, August 2012.

[8] 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.

[9] G. O. Glentis and A. Jokobssen, "Time-recursive IAA spectral estimation," *IEEE Signal Processing Letters*, vol. 18, no. 2, Feb. 2011.

[10] D.G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*, MaGraw-Hill, Inc., 2000.

[11] A. C. Kot, S. Parthasarathy, D. W. Dufts, and Y. Ding, "Statistical performance of single frequency estimation in white noise using state variable balancing and linear prediction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, pp. 1639–1642, Nov. 1987.

[12] Y. Ding and R. J. Vaccar, "Determination of data matrix dimensions for subspace-based parameter estimation algorithms," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2547 – 2550, May 1996.

[13] R. Garg, A. L. Varna, and M. Wu, "Modeling and analysis of electric network frequency signal for timestamp verification," *To appear in IEEE International Workshop on Information Forensics and Security*, Dec. 2012.

[14] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Inc., 2005.

[15] "http://www.sal.ufl.edu/newcomers/enf_data.rar," Accessed July 2012.

[16] Y. Zhang, P. Markham, T Xia, L. Chen, Y. Ye, Z. Wu, Z. Yuan, L. Wang, J. Bank, J. Burgett, R. W. Conners, and Y. Liu, "Wide-area frequency monitoring network (FNET) architecture and applications," *IEEE Transactions on Smart Grid*, vol. 1, no. 2, Sep. 2010.

[17] Web-Based Clock, "http://www.timeanddate.com," Accessed July 2012.