ENF Extraction From Digital Recordings Using Adaptive Techniques and Frequency Tracking

Ode Ojowu, Jr., *Student Member, IEEE*, Johan Karlsson, *Member, IEEE*, Jian Li, *Fellow, IEEE*, and Yilu Liu, *Fellow, IEEE*

***Abstract*—A novel forensic tool used for assessing the authenticity of digital audio recordings is known as the electric network frequency (ENF) criterion. It involves extracting the embedded power line (utility) frequency from said recordings and matching it to a known database to verify the time the recording was made, and its authenticity. In this paper, a nonparametric, adaptive, and high resolution technique, known as the time-recursive iterative adaptive approach, is presented as a tool for the extraction of the ENF from digital audio recordings. A comparison is made between this data dependent (adaptive) filter and the conventional short-time Fourier transform (STFT). Results show that the adaptive algorithm improves the ENF estimation accuracy in the presence of interference from other signals. To further enhance the ENF estimation accuracy, a frequency tracking method based on dynamic programming will be proposed. The algorithm uses the knowledge that the ENF is varying slowly with time to estimate with high accuracy the frequency present in the recording.**

***Index Terms*—Audio forensics, dynamic programming, electric network frequency (ENF) criterion, iterative adaptive approach (IAA).**

I. Introduction

THE use of digital recorders has become more prevalent in the world today due to the advancement in digital technology and the significant progress made in the field of digital signal processing (DSP). Prior to the increased use of digital recorders, forensic audio analysis relied on different techniques of audio authentication. For instance, the magnetic signatures that are left by the erase, record or play heads on the magnetic tape of analog recorders can be used to verify the authenticity of such recordings.

When it comes to digital recordings, alterations can be made very easily without leaving behind such imprints, because digital recorders produce a recording by converting sound variations to a series of numbers, making authentication of these recordings a lot more difficult [1]. The importance of being able to verify the authenticity of a recording can be seen in litigation cases [2], where digital recordings are brought forward as evidence in a trial. Therefore, more reliable methods of verifying the authenticity of digital recordings need to be researched.

The electric network frequency (ENF) criterion was proposed by Grigoras [2], [3] to address the issue of digital audio authentication. The ENF criterion is based on extracting the utility frequency or ENF from a digital audio recording and matching the extracted frequency estimate to a reference database in order to determine the authenticity and also time of the digital recording. This process is possible because, in some cases, digital recorders (even some battery powered recorders [4]), can pick up the audible sound that is generated by the oscillation of a power grid's alternating current at this frequency. The frequency of oscillation is approximately 60 Hz in the U.S., whereas in Europe it oscillates at approximately 50 Hz. The corresponding harmonics of this frequency might also be present in the digital recording.

The ENF criterion is based on two assumptions [5]. Firstly, the ENF for interconnected networks is the same at all points within the network. Secondly, the frequency varies randomly within a given interconnection, and hence, is not repeatable over a long period of time.

There are three known methods of extracting the ENF over time from a digital recording [2], [3]. They are as follows.

1) *Time/frequency domain analysis*—This method is based on computing the spectrogram of the signal and visually comparing it to the database.
2) *Frequency domain analysis*—This method is based on selecting the frequency location corresponding to the maximum amplitude of the power spectrum of segments (frames) of the data after applying a bandpass filter.
3) *Time domain analysis*—This method is based on measuring the zero crossings of the signal in the time domain after a bandpass filter has been applied to the recording.

Recently in [6], a quadratic interpolation scheme was applied to the *frequency domain analysis* method to estimate the spectral peak locations (frequencies) more accurately. This reduces the estimation error resulting from the use of a fixed grid size in the spectral estimation process.

Besides the time-domain analysis, the mentioned methods estimate the ENF based on computing the fast Fourier transform (FFT) of overlapping segments (frames) of the data known as the short-time Fourier transform (STFT), which is limited by the tradeoff between time resolution and frequency resolution [7]. Parametric methods such as the frequency selective ESPRIT, which give superior resolution compared to the FFT, can also be

Manuscript received December 02, 2011; revised February 28, 2012; accepted April 18, 2012. Date of publication May 02, 2012; date of current version July 09, 2012. This work was supported in part by the Swedish Research Council. This work made use of Engineering Research Center Shared Facilities supported by the Engineering Research Center Program of the National Science Foundation and DOE under NSF Award Number EEC-1041877 and the CURENT Industry Partnership Program. The work of Y. Liu was supported in part by Award 2009-DN-BX-K233, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dinei A. Florencio.

O. Ojowu, Jr., J. Karlsson, and J. Li, are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6130 USA (e-mail: ojowuode@ufl.edu; jkarlsson@ufl.edu; li@dsp.ufl.edu).

Y. Liu is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996 USA (e-mail: liu@utk.edu).

Digital Object Identifier 10.1109/TIFS.2012.2197391

TABLE I
NOTATIONS

$\mathbf{x}$	a vector
$\mathbf{X}$	a matrix
$\text{diag}(\mathbf{x})$	a diagonal matrix with elements of $\mathbf{x}$ on the diagonal
$(\cdot)^H$	conjugate transpose of a matrix or vector
$(\cdot)^T$	transpose of a matrix or vector
$\|\| \cdot \|\|_2$	ℓ_2 norm
$\hat{x}$	estimate of scalar x
$\triangleq$	definition

TABLE II
ABBREVIATIONS

APES	Amplitude and Phase Estimation
ENF	Electric Network Frequency
ESPRIT	Estimation of Signal Parameters by Rotational Invariance
FDR	Frequency Disturbance Recorder
FIAA	Fast Iterative Adaptive Approach
IAA	Iterative Adaptive Approach
QN-IAA	Quasi-Newton Iterative Adaptive Approach
STFT	Short-time Fourier Transform
TRIAA	Time-Recursive Iterative Adaptive Approach

used successfully to extract the ENF from one frame to another. However, in the presence of significant interference within a given frame, the parametric methods yield poor frequency estimates because of their sensitivity to an assumed data model.

This paper focuses on two methods of extraction. The first builds upon the *frequency domain analysis* with quadratic interpolation. However, in place of the FFT, the spectrum is estimated for each segment of the data using a nonparametric and high resolution adaptive algorithm known as the iterative adaptive approach (IAA) [8]. In the presence of interfering signals with frequencies within the range of values the ENF can take on, IAA yields more accurate estimates of the ENF compared to the FFT as a result of the improved spectral resolution and interference suppression capability. The second method involves applying a frequency tracking algorithm based on discrete dynamic programming [9], which takes into account the slowly varying nature of the ENF over time. This tracking algorithm is necessary because, in some frames of the data, the maximum spectral peak might correspond to an interference signal rather than the network frequency signal even within the acceptable ENF limits. The ENF is then estimated inaccurately, which can result in a false diagnosis that the recording in question has been edited.

It is worthwhile to point out that, in order for the proposed methods to work, the ENF must be embedded in the recording, which is not always the case especially in some battery operated recorders [4]. This is certainly a drawback of using the ENF criterion for digital authentication. However, if the ENF is embedded in a digital recording, more reliable methods of extraction need to be sought.

Extraction can also be carried out using the harmonics of the ENF signal for the frequency estimation process. In some cases, the harmonics may give better estimates because of a higher signal-to-interference-and-noise ratio compared to the fundamental frequency.

The remaining sections of this paper are organized as follows. In Section II, the network characteristics and the network frequency database are described. In Section III, the IAA and TRIAA algorithms are described along with the frequency tracking algorithm for ENF extraction. In Section IV, the experimental results based on a set of digital audio recordings are presented. Finally, Section V contains the conclusions drawn from the results.

Notation: Boldface uppercase and lowercase letters are used to denote matrices and vectors, respectively. See Table I for more details on notation.

Abbreviations: The abbreviations are presented for easy reference in Table II.

II. NETWORK FREQUENCY CHARACTERISTICS AND DATABASE

The frequency at which alternating current is distributed to various customers from power stations corresponds to the utility frequency or ENF. For European and most Asian countries the value of this frequency is 50 Hz, while the value is 60 Hz in North America and several countries in South America. Japan uses both frequencies (50 and 60 Hz) for electricity distribution. This frequency is determined by the speed of rotation of the turbines used to drive the generators at the various power plants [11]. Naturally, the rotation speed is not constant and varies within a certain limit (approximately ± 0.05 Hz) depending on the amount of load connected to the network and amount of power generated at a given time. Experiments carried out in some European countries [2], [12] have shown that this frequency variation is random and unique within specific geographic locations. This uniqueness in frequency variation within a region, coupled with the fact that network frequency is not repeatable over a long period of time, is what makes the aforementioned ENF criterion possible.

A database of the network frequency is needed in order to match the extracted ENF from a recording for verification. In [2], such a database is created by connecting the sound card of a computer to a transformer which is then connected directly to an ac power outlet. The database currently being built in North America involves deploying several sensors termed frequency disturbance recorders (FDRs), which perform accurate ENF measurements, up to about ± 0.0005 Hz. The measured data collected by the FDRs is transmitted over the internet to servers, where it can be analyzed and stored in a system termed the information management system (IMS) [13]. This collection forms the frequency monitoring network (FNET).

There are two major interconnections in North America and three minor interconnections. These regions have unsynchronized networks (frequency and phase) and are therefore connected via high voltage direct current lines (HVDC) [14]. The Eastern and Western interconnections form the major interconnections, while the Quebec, Texas and Alaska interconnections form the minor. The Alaska interconnection is isolated, in the sense that it is not connected to any of the other interconnections. It is therefore generally not considered to be part of the North American grid. Fig. 1 shows the distribution of the FDRs

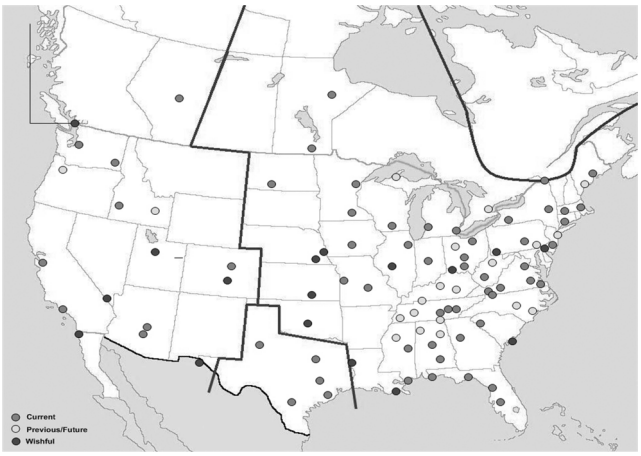

Fig. 1. FDR distribution in North America [10].

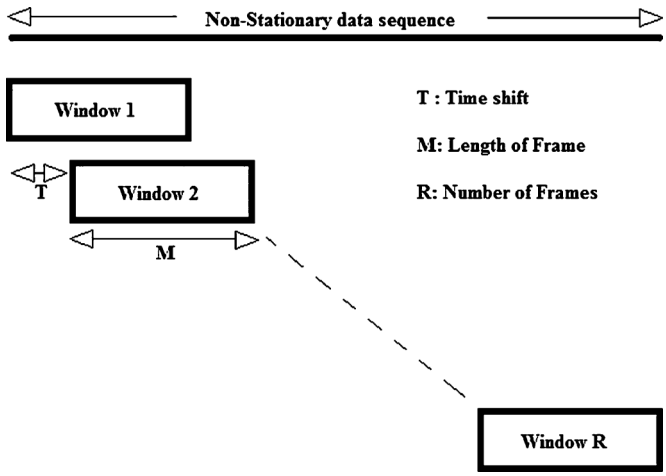

Fig. 2. Segmentation of data for STFT.

in Western, Eastern, Quebec and Texas Interconnections. Frequency measurements collected by the FDRs in these interconnections show that the frequency pattern is different at a given time from one interconnection to another. However, the frequency pattern is unique at different locations within each interconnection [10]. The FNET system, therefore, provides a viable ENF database.

III. Extraction Algorithms

A. Frequency Domain Analysis (STFT) [2]

Due to the fact that the ENF varies with time, the extraction process involves analyzing a nonstationary data sequence. STFT is a common method for time-frequency analysis of signals. This analysis assumes the signal of interest is stationary within short time windows (frames); the FFT of the signal is then computed for each frame. The *frequency domain analysis* [2] method of extraction is based on this idea.

The process involves resampling the audio signal to a lower sampling rate, to reduce the computational complexity of the analysis. A bandpass filter with a narrow bandwidth is applied to the signal with center frequency 50/60 Hz as a preprocessing step. The rest of the analysis is described as follows. Let

$$\mathbf{z} = [z_0, z_1 \ldots z_{N-1}]^T \tag{1}$$

denote the resampled and filtered discrete-time signal. This signal is then split into R overlapping frames as shown in Fig. 2, with each frame having length M and a shift from frame to frame of length T. Using the *frequency domain analysis* method, the ENF of the rth frame is estimated by finding the frequency that maximizes the spectrum of each frame which is computed using the FFT-based periodogram.

In order to get a more accurate estimate of the frequency, quadratic interpolation is used [6], [15]. This interpolation scheme involves fitting a quadratic model of the form

$$\log \hat{\phi}(\omega) = m\left(\omega - \omega_{k_{\max}} - \Delta\right)^2 + c \tag{2}$$

around the frequency point that maximizes the power spectrum

$$\omega_{k_{\max}} = \arg\max_{\omega_k} \phi_r(\omega_k) \tag{3}$$

where $\omega_k = 2\pi k/K, k = 0, 1, \ldots, K-1$ corresponds to the frequency grid point of a frequency grid with size K, and $\phi_r(\omega_k)$ is power spectrum of the rth frame.

The value of ω that maximizes the model (2) is taken as the estimated peak of the spectrum. This value is determined by fitting the model to the highest sample of the power spectrum and the two adjacent points with corresponding frequencies $(\omega_{k_{\max}-1}, \omega_{k_{\max}}, \omega_{k_{\max}+1})$. This value of ω that maximizes the model is

$$\omega = \omega_{k_{\max}} + \Delta \tag{4}$$

where

$$\Delta = \frac{1}{2}\frac{\beta_{-1} - \beta_1}{\beta_{-1} - 2\beta_0 + \beta_1}\left(\omega_{k_{\max}+1} - \omega_{k_{\max}}\right) \tag{5}$$

$$\beta_\ell \triangleq \log \phi_r\left(\omega_{k_{\max}+\ell}\right), \quad \ell = -1, 0, 1. \tag{6}$$

The corresponding frequency estimate of the rth frame in Hz is given by

$$\hat{f}(r) = 2\pi\left(\omega_{k_{\max}} + \Delta\right) F_s \tag{7}$$

where F_s is the sampling frequency (in Hertz) of the signal.

The use of STFT will result in a tradeoff between frequency resolution and time resolution. For a given frame length, this tradeoff can be optimized by applying a rectangular window to each frame, which will provide the best spectral resolution at a cost of higher side lobes compared to other spectral windows.

In order to get improved spectral resolution over FFT, one has to resort to using parametric methods or data-dependent (adaptive) nonparametric methods for spectral estimation. Parametric methods, on the one hand, are not robust against data model errors. On the other hand, nonparametric adaptive methods are more robust, since they do not assume a specific parametric data model. Well-known adaptive methods include the Capon

algorithm and the amplitude and phase estimation (APES) algorithm. These algorithms also provide higher resolution and lower side lobes than the periodogram. However, these methods are inadequate because they require multiple realizations (snapshots) of the random signal, which is not the case with the current data, as only one snapshot is available for frequency estimation. Spatial smoothing (segmenting and spectral averaging of the data) can be used to improve the spectral estimates of the Capon and APES algorithms in the one-snapshot case; but the cost of doing this will be a degradation in the spectral resolution, which is not desirable. The wavelet transform is also a common tool for time-frequency analysis. Contrary to the STFT, which uses a fixed window size, the wavelet transform uses short windows at high frequencies and longer windows at low frequencies. The wavelet transform is therefore not suitable for our problem because we are interested only in a small range of frequencies.

IAA is a nonparametric data-dependent algorithm based on weighted least squares (WLS), originally presented in [8] for direction of arrival (DOA) estimation in array processing. The IAA algorithm is capable of yielding high resolution and low side lobes even in the case of a single snapshot [8], hence making it suitable for estimating the ENF in the presence of interferences.

B. IAA and TRIAA

The ENF can be extracted with high accuracy in the presence of interference using the IAA algorithm for a given frame. The proposed ENF extraction process follows (2)–(7), with the FFT spectral estimate ϕ_r replaced by the IAA spectral estimate. The IAA and TRIAA [16] used for spectral estimation of nonstationary data will be discussed in this section.

The spectral estimation problem can be setup as follows. Let $\mathbf{y} = [y_0, y_1 \dots y_{M-1}]^T$ denote a uniformly sampled stationary data sequence and $\mathbf{A} = [\mathbf{a}(\omega_0), \mathbf{a}(\omega_1) \dots \mathbf{a}(\omega_{K-1})]$, where $\mathbf{a}(\omega_k) = [1, e^{j\omega_k}, \dots, e^{(M-1)j\omega_k}]^T$ corresponds to a steering (frequency) vector, and $\omega_k = 2\pi k/K, k = 0, 1, \dots, K-1$ corresponds to a frequency grid point of a frequency grid with size K. Also let $\boldsymbol{\alpha} = [\alpha(\omega_0), \alpha(\omega_1), \dots, \alpha(\omega_{K-1})]^T$, with $\alpha(\omega_k)$ denoting the complex spectral estimates of $\mathbf{y}$ at ω_k. The following data model can be formulated:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\alpha} \tag{8}$$

where the noise contributions of $\mathbf{y}$ are taken into account implicitly [8].

The IAA algorithm solves for the spectral estimates $\boldsymbol{\alpha}$ by minimizing the following quadratic cost function in (9) using weighted least squares (WLS):

$$\|\mathbf{y} - \mathbf{a}(\omega_k)\alpha(\omega_k)\|^2_{\mathbf{Q}_{-1}}(\omega_k) \tag{9}$$

where $\|\mathbf{x}\|^2_{\mathbf{Q}^{-1}}(\omega_k) \triangleq \mathbf{x}^H\mathbf{Q}^{-1}(\omega_k)\mathbf{x}$

$$\mathbf{Q}(\omega_k) = \mathbf{R} - p_k\mathbf{a}(\omega_k)\mathbf{a}^H(\omega_k) \tag{10}$$

$$\mathbf{R} = \mathbf{APA}^H \tag{11}$$

and $\mathbf{P} \triangleq \operatorname{diag}[p_0, p_1, \dots p_{K-1}]$, with p_k for $k = 0, \dots, K-1$, denoting the power estimate at each frequency grid point, given by $|\alpha(\omega_k)|^2$. $\mathbf{R}$[1] is the covariance matrix of the data and $\mathbf{Q}(\omega_k)$ is the covariance matrix of the interference and noise, where interference refers to all the signals at frequency grid points other than the current grid point of interest ω_k. Minimizing the cost function in (9) with respect to the $\alpha(\omega_k)$ for $k = 0, \dots, K-1$ gives the following solution:

$$\hat{\alpha}(\omega_k) = \frac{\mathbf{a}^H(\omega_k)\mathbf{Q}^{-1}(\omega_k)\mathbf{y}}{\mathbf{a}^H(\omega_k)\mathbf{Q}^{-1}(\omega_k)\mathbf{a}(\omega_k)}, \quad k = 0, 1, \dots, K-1. \tag{12}$$

The solution in (12) can be rewritten as

$$\hat{\alpha}(\omega_k) = \frac{\mathbf{a}^H(\omega_k)\mathbf{R}^{-1}\mathbf{y}}{\mathbf{a}^H(\omega_k)\mathbf{R}^{-1}\mathbf{a}(\omega_k)}, \quad k = 0, 1, \dots, K-1 \tag{13}$$

using the Woodbury matrix identity[2] and (10). This prevents the computation of the interference covariance matrix $\mathbf{Q}^{-1}(\omega_k)$ for each frequency grid point. Note that the computation of $\mathbf{R}^{-1}$ requires the knowledge of $\alpha(\omega_k)$ and vice versa. Hence this algorithm is solved in an iterative manner, with the estimate of $\boldsymbol{\alpha}$ initialized using the FFT. This iterative algorithm takes about 10 to 15 iterations to converge based on experimental and numerical results.

Note also that without accounting for the interference from other frequency grid points (without weighting), minimizing the cost function in (9) for $K = M$ gives the discrete Fourier transform (DFT) of the signal

$$\hat{\alpha}(\omega_k) = \frac{\mathbf{a}^H(\omega_k)\mathbf{y}}{M}, \quad k = 0, 1, \dots, M-1. \tag{14}$$

The IAA algorithm described is used for spectral estimation of stationary data. Analogous to the STFT, the spectral content of a nonstationary data sequence, such as (1), can be estimated using the TRIAA [16]. The signal is split into overlapping frames similar to Fig. 2 and the IAA spectral estimate is computed for each frame. However, to reduce the computational complexity, each subsequent frame after the first frame is initialized with the spectral estimate of the previous frame instead of the FFT-based periodogram as described in the IAA algorithm. The resulting algorithm yields better spectral resolution and lower side lobes than the STFT.

There is still a significant increase in the computational complexity when using the TRIAA algorithm compared to using STFT for spectral estimation. This computational complexity is reduced slightly by reducing the number of iterations in subsequent frames for the TRIAA. This is because convergence of the estimated spectrum will occur in fewer iterations given the current frame is initialized by the spectral estimate of the previous frame. When the dataset is significantly large, the use of this algorithm is still impractical. The bottleneck of the TRIAA algorithm is in the computation of the denominator in (13) for each frame.

In [18] and [19] the Toeplitz structure of the covariance matrix $\mathbf{R}$ is exploited and the computation of $\mathbf{R}^{-1}$ is performed

[1] $\mathbf{R} = \mathbf{APA}^H + \sigma^2\mathbf{I}$ for ill-conditioned matrices [17].

[2] Matrix inversion lemma.

using the Gohberg-Semencul (GS) factorization of this matrix [7]. Moreover, the denominator is obtained via evaluating a polynomial. This reduces the computational complexity of the denominator in (13) (which is the bottleneck of the IAA algorithm) from $\mathcal{O}(M^2K)$ to $\mathcal{O}(M^2)$ floating point operations (flops) [18] for a given frame, without a loss in performance. The algorithm is termed the Fast IAA (FIAA), which is a significant improvement but still computationally expensive for large datasets. The computational complexity of IAA and FIAA are $\mathcal{O}(M^2K)$ and $\mathcal{O}(M^2+K\log K)$, respectively, where M is the data length and K is the grid size, with $K \gg M$.

An approximate algorithm to the IAA algorithm with significantly faster computational time is described in [20] and referred to as the Quasi-Newton IAA (QN-IAA). The QN-IAA algorithm estimates the covariance matrix as if it were from a low-order (L) autoregressive (AR) process, where $L \ll M$ with M being the data (frame) length. The inversion of the lower order covariance matrix $\mathbf{Q} \in \mathbb{C}^{L\times L}$ is carried out in place of $\mathbf{R} \in \mathbb{C}^{M\times M}$, yielding an approximate solution to the IAA spectral estimate (13) with significant reduction in the computational complexity and just a slight degradation in the resolution. The computational complexity of this algorithm is $\mathcal{O}(L^2 + K\log K)$.

The FIAA or QN-IAA can be used in a time-recursive manner for nonstationary data as is the case with the ENF signal. This algorithm reduces the tradeoff between frequency resolution and time-resolution for a given frame length compared to the FFT-based periodogram during the ENF extraction process. The extraction process is the same as the *frequency domain analysis* (2)–(7) with ϕ_r replaced by either of the aforementioned algorithms.

However, even if a good algorithm is used for frequency estimation based on (7), specific frames might be corrupted by interference signals with frequency components within the ENF limits. This could lead to errors in frequency estimation, if the frequency location corresponding to the maximum value of the estimated spectra belongs to an interference signal. A robust method of tracking the ENF that exploits the slowly varying nature of this frequency is needed. The next section describes the proposed frequency tracking algorithm.

C. Frequency Tracking

A method of estimating the ENF by tracking it from one frame to another is formulated here from a mathematical point of view. The proposed method uses discrete dynamic programming [9] to find a minimum cost path. A cost function as shown in this section is selected which takes into account the slowly varying nature of the actual network frequency. This cost function penalizes significant jumps in frequency from frame to frame and the corresponding path is used to estimate the ENF.

This algorithm involves finding the peak locations from the spectrum of each frame and assigning costs based on the difference between a peak location in one frame and a peak location in another frame. The magnitude of the assigned cost is related to the difference in the frequency from one frame to another. The minimum cost path from the first frame to the last frame is computed to estimate the ENF.

To estimate the number of relevant peaks (sinusoids) in a given frame, a model order selection tool known as the Bayesian Information Criterion (BIC) is used. The BIC for complex sinusoids in noise is given by (refer to [7] and [21] for a full derivation)

$$\mathrm{BIC}(n_r) = M\ln\left(\left\|\mathbf{y}-\sum_{k=1}^{2n_r}\mathbf{a}(\omega_k)\hat{\alpha}(\omega_k)\right\|^2\right) + 5(2n_r)\ln M. \quad (15)$$

The number of peaks (real sinusoids) n_r is estimated as the minimizing argument of the above BIC criterion. The first term in (15) is a least squares data fitting term, which decreases as the number of estimated peaks n_r increases, where the second term is a penalty term that prevents "overfitting" of the data model. Once the n_r largest peaks and corresponding locations are determined in each frame, the frequency tracking problem is formulated and solved as follows.

Assume that for a given frame r, a set of estimated peak locations (frequencies) is denoted by $\Lambda_r = \{P_{r1}, P_{r2}, \ldots P_{rn_r}\}$. We would like to find a path $\{f_r\}_{r=1}^{R}$ such that $f_r \in \Lambda_r$ and where the difference $f_r - f_{r-1}$ is as small as possible for $r = 1, 2, \ldots, R$. This set corresponds to the estimated ENF over all frames and can be obtained as the minimizing argument in the following optimization problem:

$$J = \min_{\substack{f_r\in\Lambda_r \\ r=1,\ldots,R}} \sum_{r=2}^{R}(f_r - f_{r-1})^2. \quad (16)$$

Calculating this cost using an exhaustive search is impractical. However, using dynamic programming [9] the path that minimizes this cost can be computed recursively and efficiently by minimizing the cost from a given frame $j < R$, to the last frame, denoted by $J(j, f_j)$

$$J(j,f_j) = \min_{\substack{f_r\in\Lambda_r \\ r=j+1,\ldots,R}} \sum_{r=j+1}^{R}(f_r - f_{r-1})^2,\ f_j \in \Lambda_j. \quad (17)$$

This optimal cost satisfies the recursive equation

$$J(j,f_j) = \min_{\substack{f_{j+1} \\ \in\Lambda_{j+1}}} \left\{(f_{j+1}-f_j)^2 + J(j+1, f_{j+1})\right\}, \quad f_j \in \Lambda_j \quad (18)$$

which can be calculated for $j = R-1, R-2, \ldots, 1$, with the initialization, $J(R, f_R) = 0,\ f_N \in \Lambda_N$. Note that

$$J = \min_{f_1\in\Lambda_1} J(1, f_1),\ f_N \in \Lambda_N \quad (19)$$

is the cost from the first frame to the last frame R and the set $\{f_r\}_{r=1}^{R}$ that minimizes this cost function corresponds to the extracted ENF signal as mentioned previously. Dynamic programming has a computational complexity of $\mathcal{O}(R\Lambda_{\max}^2)$, where R corresponds to the total number of frames and $\Lambda_{\max}$

TABLE III
PARAMETERS FOR THE EXPERIMENT

PARAMETERS	*'Data1'*	*'Data2'*
T (Time Shift)	1s	1s
M (Length of Frame)	20s	33s
R (Number of Frames)	1800	1800

is the number of spectral peaks in the frame with the maximum number of peaks.

D. Matching Extracted ENF to Database

Once the ENF signal has been extracted, a method of matching the estimated signal to the database signal is required. The goal is to find the location/time within the database that is similar in pattern to the extracted ENF. In [6], a method based on minimizing the squared error between the ENF and database is used for automated matching. A method of correlation matching proposed in [22] for short digital recordings (10–15 min) is used in place of this MSE method. The process of correlation matching is described as follows. Assume that $\mathbf{f} = [f_1, f_2, \ldots, f_R]$ is the extracted ENF signal and $\mathbf{d} = [d_1, d_2, \ldots, d_L]$ corresponds to the database signal with $L > R$. The matching process requires finding $l_{\max}$ such that

$$l_{\max} = \arg\max_{l} c(l), \quad l = 1, 2, \ldots, L - R \tag{20}$$

where $c(l)$ is the correlation coefficient between $\mathbf{f}$ and the vector $[d_l, d_{l+1}, \ldots, d_{l+R-1}]$.

An important point to make is that the maximum correlation coefficient $c(l_{\max})$ is used here only for matching the estimated ENF to the database and comparing the accuracy (reliability) of the different algorithms presented. Once a match has been made, determining locations of edits to a recording should be based on the differences between the ENF estimate and the database.

IV. EXPERIMENTAL RESULTS

The algorithms presented in the previous section are applied to two different digital audio datasets referred to as "*Data1*" and "*Data2*". The two datasets are recorded simultaneously and, therefore, should contain the same ENF pattern over time. The first data set (*Data1*) is acquired by connecting an electric outlet via a voltage divider directly to the internal sound card of a desktop computer, resulting in an ENF signal with a rather high signal-to-interference-and-noise ratio. On the other hand, the second dataset (*Data2*) is an actual speech recording played from a speaker and picked up by the internal microphone of a laptop computer.

Each of these recordings are originally sampled at 44.1 kHz at a bit rate of 16 bits per sample. Each dataset is resampled to 441 Hz, hence keeping only the fundamental frequency (1st harmonic) and the two higher harmonics of the ENF. A bandpass filter with a narrow bandwidth around the network frequency is applied to the data to eliminate as much interference as possible without distorting the ENF signal. Based on Fig. 2, each data set is split using the values shown in Table III. This setup results in an ENF estimate every second for a total of 30 min for each dataset.

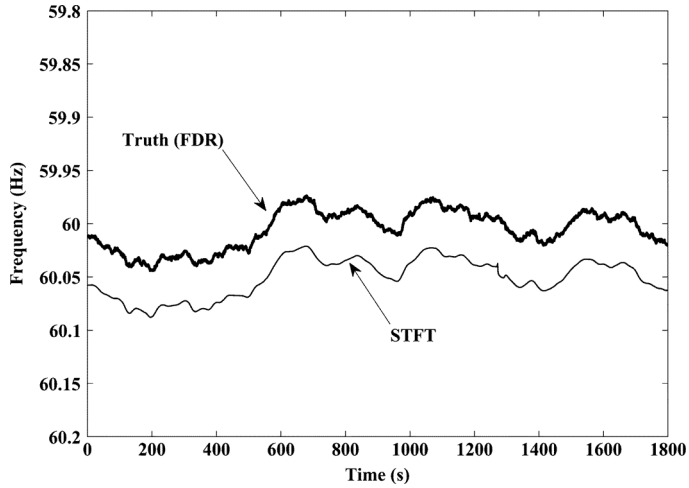

Fig. 3. Matching extracted ENF to database (*Data1*—scaled to 60 Hz).

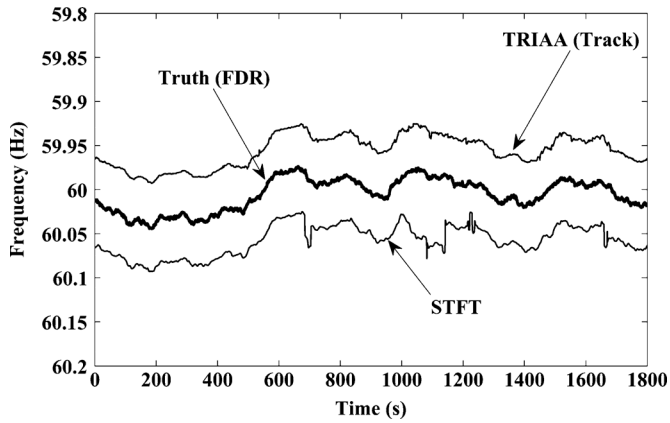

Fig. 4. Matching extracted ENF to database (*Data2*—scaled to 60 Hz).

An increase in the frame length improves the signal-to-noise ratio of the signal [6] and the spectral resolution at the cost of lower time resolution. Therefore, a larger frame length is used for *Data2* which has a weak ENF signal compared to *Data1* which has a strong ENF signal.

Fig. 3 shows the extracted ENF signal (shifted by 0.05 Hz for illustration purposes) from *Data1*, matched with the truth obtained from the FDRs, when the data set has not been altered in any form [using STFT and (7)]. Fig. 4 shows the extracted ENF using the STFT-based method and our proposed method (also shifted for comparison purposes). Tables IV and VI give the maximum correlation coefficient $c(l_{\max})$ of the various methods for *Data1* and *Data2*, respectively, also when the signals have not been altered. The maximum correlation coefficient values are used to compare the accuracy of the algorithms and hence determine which is more reliable for ENF estimation. We have also included similar MSE (actually standard deviation) analysis in Tables V and VII for the datasets, where the MSE is computed by averaging the squared difference between the True ENF and the estimated ENF. It is important to

TABLE IV
CORRELATION COEFFICIENTS OF ALGORITHMS (*DATA1*)

	Algorithm				
Harmonic	STFT	STFT(Track)	TRIAA	TRIAA(Track)	F-ESPRIT
60 Hz	0.9912	0.9917	0.9895	0.9900	0.9800
120 Hz	0.9911	0.9949	0.9902	0.9946	0.9470
180 Hz	0.9968	0.9968	0.9961	0.9961	0.9962

TABLE V
STANDARD DEVIATION OF ERROR FOR ALGORITHMS (*DATA1*)

	Algorithm				
Harmonic	STFT	STFT(Track)	TRIAA	TRIAA(Track)	F-ESPRIT
60 Hz	$2.772e^{-3}$	$2.650e^{-3}$	$3.032e^{-3}$	$2.919e^{-3}$	$5.364e^{-3}$
120 Hz	$2.774e^{-3}$	$2.145e^{-3}$	$2.822e^{-3}$	$2.198e^{-3}$	$6.570e^{-3}$
180 Hz	$1.900e^{-3}$	$1.851e^{-3}$	$1.999e^{-3}$	$1.999e^{-3}$	$2.830e^{-3}$

point out that the estimated ENF can sometimes have a constant offset [12], [22]. Therefore, the correlation is the preferred method for accuracy measure. The datasets used for this experiment do not have such an offset. They have also been made available at http://www.sal.ufl.edu/download.html.

A. Data1 Analysis

Fig. 3 shows the extracted harmonic (180 Hz) of the ENF signal scaled to 60 Hz and matched [using the location corresponding to the maximum correlation (20)] to the actual database frequency obtained from the FDRs. For each of the algorithms used, the third harmonic gave the most accurate results for this dataset as shown in Table IV. This is because for a fixed grid size, the estimation error when using the third harmonic is reduced by a factor of three compared to the fundamental frequency. Harmonics with frequencies higher than 180 Hz can be used for the estimation process at a cost of increased computational complexity due to the increased sampling rate. Also from Table IV, it can be seen that each of the STFT and TRIAA algorithms produce accurate estimates of the ENF using (7) because of the rather strong ENF signal. The signal at the second harmonic is weak relative to the first and third harmonics, and in a few frames the estimate was inaccurate. However, the frequency tracking algorithm mitigated these inaccuracies successfully by tracking the correct spectral peaks.

The parametric method, frequency selective (F-ESPRIT) [7], [23] also yields accurate estimates of the ENF for *Data1* when the signal model assumes there is only one sinusoid per frame. However, this method and other parametric methods are not appropriate for ENF estimation in the presence of interference, because they are sensitive to model assumptions.

For this dataset, the STFT yields slightly better results, compared to the adaptive method (TRIAA). This can be explained by the fact that the periodogram is optimal for estimating spectral lines (sinusoids) in the presence of white noise when they are well resolved [7]. However, when there are interfering signals present, the poor resolution of the periodogram will yield inaccurate estimates as is the case with *Data2*, a typical digital recording.

TABLE VI
CORRELATION COEFFICIENTS OF ALGORITHMS (*DATA2*)

	Algorithm				
Harmonic	STFT	STFT(Track)	TRIAA	TRIAA(Track)	F-ESPRIT
120 Hz	0.9125	0.9857	0.9305	0.9907	0.8446

TABLE VII
STANDARD DEVIATION OF ERROR FOR ALGORITHMS (*DATA2*)

	Algorithm				
Harmonic	STFT	STFT(Track)	TRIAA	TRIAA(Track)	F-ESPRIT
120 Hz	$7.948e^{-3}$	$3.369e^{-3}$	$7.225e^{-3}$	$2.914e^{-3}$	$1.086e^{-2}$

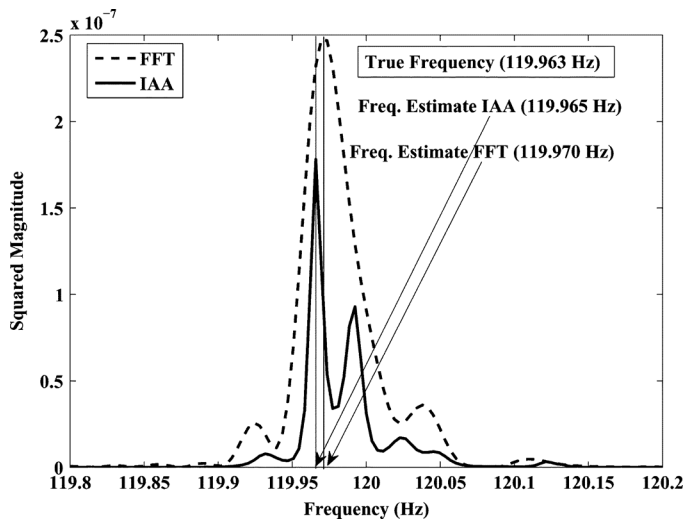

Fig. 5. Power spectrum of one frame (*Data2*): poor resolution of FFT.

B. Data2 Analysis

For *Data2*, the second harmonic (120 Hz) is used to estimate the ENF, because the first and third harmonics are too weak to be used for estimation. Table VI shows the maximum correlation coefficient values for the STFT and TRIAA using (7), the frequency tracking algorithm using the spectral peaks of the FFT and IAA and the parametric method (F-ESPRIT) with one assumed sinusoid. The ENF estimation accuracy is improved using the adaptive method (IAA) because of improved spectral resolution for several frames. Fig. 5 shows a comparison of the spectrum of one frame of the *Data2*, where the poor frequency resolution of the FFT results in a relatively poor estimate of the network frequency compared to the IAA algorithm.

Fig. 4 shows this extracted ENF harmonic using the STFT and (7) matched with the database. From this figure, there are several frames where the ENF is estimated inaccurately, due to the fact that the frequency corresponding to the maximum spectral peak for those frames do not correspond to the ENF. This can occur if there is another signal present with frequency within the limits of the acceptable range of the ENF as illustrated in Fig. 6. This figure shows that for both spectral estimation techniques used (IAA, FFT) the ENF harmonic estimate using (7) will be 120 Hz, whereas the true frequency is approximately 119.95 Hz.

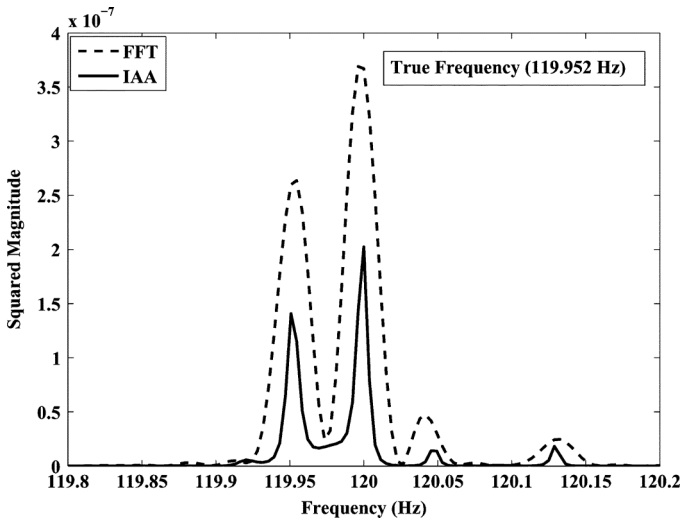

Fig. 6. Power spectrum of one frame (*Data2*): strong interference signal.

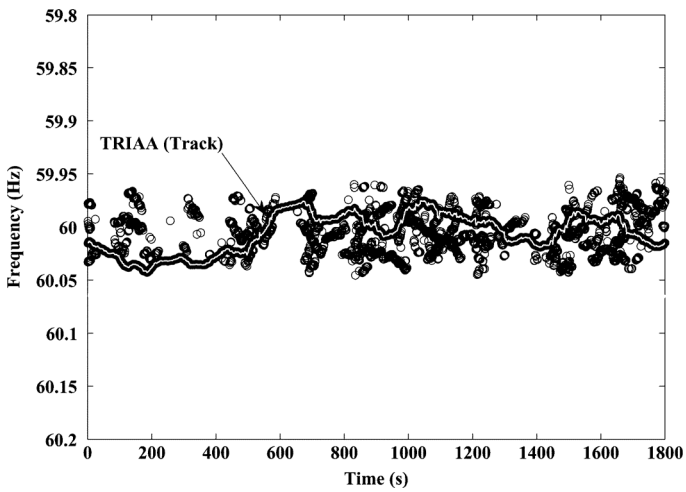

Fig. 7. Extracted ENF via frequency tracking (*Data2*—scaled to 60 Hz).

This problem can be rectified using our dynamic programming-based frequency tracking the algorithm presented.

Fig. 7 shows the spectral peak locations computed using the TRIAA and the corresponding ENF estimate using dynamic programming. The estimate of the network frequency using this tracking algorithm is then matched to the database in Fig. 4, which provides a better match when compared to using (7), which can also be seen in this figure, Fig. 8 (absolute error) and also from Table VI.

A few important points to make are that the frequency tracking algorithm uses the peak locations for each frame estimated either by the adaptive algorithm (IAA) or the FFT. The results show that the estimated ENF is more accurate when the peak locations of IAA are used. This is as a result of the inaccurate estimates in some frames caused by the poor resolution of using FFT. Also, all the numbers presented can be improved upon slightly by using the entire dataset (44.1 kHz) for analysis. For example, the STFT maximum correlation of 0.9125 will be improved to 0.9158 without resampling, which may not be worth the increased computational complexity.

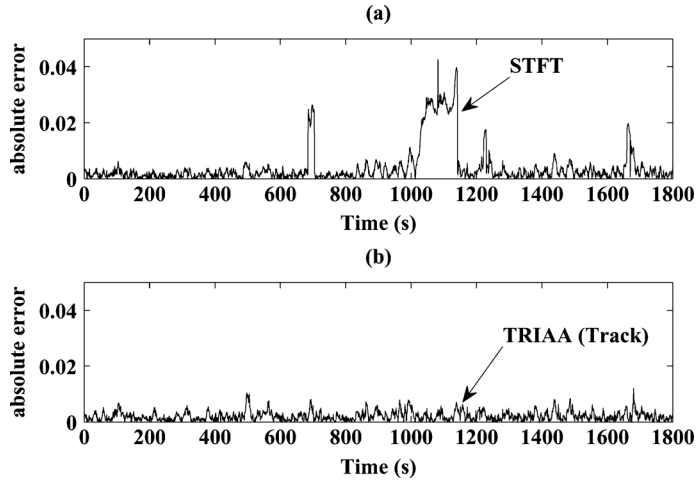

Fig. 8. Absolute error of algorithms (*Data2*): (a) STFT and (b) TRIAA (Track).

V. Conclusion

When it comes to digital audio verification, the reliability of the method used for authentication cannot be overemphasized. This paper demonstrates a reliable method of extracting the network frequency from a digital recording when the ENF cannot be extracted from some of the frames using the FFT-based periodogram either because of poor spectral resolution or a stronger interference signal within said frame. These problems were solved by using an iterative adaptive method (IAA), which provides better spectral resolution than the FFT-based approach. Also, a frequency tracking method based on dynamic programming was used for accurate extraction of the ENF even in the presence of a strong interference signals within ENF limits.

From the results presented, the FFT gives slightly better estimates of the network frequency when the signal-to-interference-plus-ratio is very high as is the case with the first dataset. However, in most digital recordings, there will be significant interferences from the recorded speech signals and other surrounding sounds that could lead to poor estimation performance using the FFT due to its poor resolution and high side lobe problems. As the results have shown, the adaptive techniques and frequency tracking method should be adopted for ENF estimation, especially in challenging environments.

Acknowledgment

The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

References

[1] R. C. Maher, "Audio forensic examination—Authenticity, enhancement, and interpretation," *IEEE Signal Processing Mag.*, vol. 26, pp. 84–94, Mar. 2009.

[2] C. Grigoras, "Digital audio recording analysis: The electric network Frequency criterion," *Int. J. Speech Language Law*, vol. 12, no. 1, pp. 63–76, 2005.

[3] C. Grigoras, "Applications of ENF criterion in forensic audio, video and telecommunications analysis," *Forensic Sci. Int.*, vol. 167, pp. 136–176, 2007.

[4] E. B. Brixen, "ENF—Quantification of the magnetic field," in *Proc. AES 33rd Conf., Audio Forensics—Theory and Practice*, Denver, CO, Jun. 2008.
[5] E. B. Brixen, "Techniques for the authentication of digital audio recordings," in *Proc. AES 122nd Conv.*, Vienna, Austria, 2007.
[6] A. J. Cooper, "The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings—An automated approach," in *Proc. AES 33rd Conf., Audio Forensics—Theory and Practice*, Denver, CO, Jun. 2008, pp. 1–6.
[7] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice-Hall, 2005.
[8] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Trans. Aerospace Electron. Syst.*, vol. 46, no. 1, pp. 425–443, Jan. 2010.
[9] U. Jönsson, C. Trygger, and P. Ögren, "Lecture Notes on Optimal Control: Optimization and system theory," unpublished.
[10] Liu, Z. Yuan, P. N. Markham, R. Conners, and Y. Liu, "Wide-area frequency as a criterion for digital audio recording authentication," in *Proc. IEEE Power Energy Soc. General Meeting*, Jul. 2011, pp. 1–7.
[11] D. Rodríguez, J. Apolinário, and L. Biscainho, "Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," *IEEE Trans. Inform. Forensics Security*, vol. 5, no. 9, pp. 534–543, Sep. 2010.
[12] M. Kajstura, A. Trawinska, and J. Hebenstreit, "Application of the electrical network frequency (ENF) criterion—A case of a digital recording," *Forensic Sci. Int.*, vol. 155, pp. 165–171, 2005.
[13] Y. Liu, "A US-wide power systems frequency monitoring network," in *Proc. IEEE Power Systems Conf. Expo.*, Atlanta, GA, Oct. 29–Nov. 1 2006, pp. 159–166.
[14] N. G. Hingorani, "High-voltage DC transmission—A power electronics workhorse," *IEEE Spectrum*, vol. 33, pp. 63–72, Apr. 1996.
[15] J. O. Smith and X. Serra, "PARSHL an analysis/synthesis program for non-harmonic sounds based on sinusoidal representation," in *Proc. Int. Computer Music Conf.*, San Francisco, CA, 2004.
[16] G. Glentis and A. Jakobsson, "Time-recursive IAA spectral estimation," *IEEE Signal Processing Lett.*, vol. 18, pp. 111–114, Feb. 2011.
[17] W. Roberts, P. Stoica, J. Li, T. Yardibi, and F. Sadjadi, "Iterative adaptive approaches to MIMO radar imaging," *IEEE J. Select. Topics Signal Process.*, vol. 4, pp. 5–20, Feb. 2010.
[18] M. Xue, L. Xu, and J. Li, "IAA spectral estimation: Fast implementation using the Gohberg-Semencul factorization," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3251–3261, Jul. 2011.
[19] G. Glentis and A. Jakobsson, "Efficient implementation of iterative adaptive approach spectral estimation techniques," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4154–4167, Sep. 2011.
[20] G. Glentis and A. Jakobsson, "Superfast approximative implementation of the IAA spectral estimate," *IEEE Trans. Signal Process.*, to be published.
[21] P. Stoica, J. Li, and H. He, "Spectral analysis of nonuniformly sampled data: A new approach versus the periodogram," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 843–858, Mar. 2009.
[22] M. Huijbregtse and Z. Geradts, "Using the ENF criterion for determining the time of recording for short digital audio recordings," in *Proc. 3rd Int. Workshop Computational Forensics, IWCF'09*, 2009, vol. 1, pp. 116–124.
[23] J. Gunarsson and T. McKelvey, "High SNR performance analysis of F-ESPRIT," in *Conf. Rec. 38th Asilomar Conf. Signals, Systems Computers*, Nov. 2004, vol. 1, pp. 1003–1007.

Ode Ojowu, Jr. (S'11) was born in Zaria, Nigeria, in 1984. He received the B.Sc. and M.Sc. degrees in electrical engineering from Washington University, St. Louis, MO, in 2007. He is currently pursuing a Ph.D. degree with the Department of Electrical Engineering at the University of Florida, Gainesville.

His primary research interests are in the areas of spectral estimation and array signal processing.

Johan Karlsson (S'06–M'09) was born in Stockholm, Sweden, in 1979. He received the M.S. degree in engineering physics and the Ph.D. degree from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2003 in 2008, respectively. He spent the academic year 2000 to 2001 as an exchange student at Washington University, Saint Louis, MO, and did his master thesis at the University of Minnesota, Minneapolis.

In Fall 2003, he was a graduate student at the Division of Optimization and Systems Theory, KTH. From 2009 to 2011, he was with Sirius International, Stockholm, Sweden. He is currently working as a Postdoctoral Research Associate in the Department of Computer and Electrical Engineering, University of Florida, Gainesville. His research interests includes fundamental limitations in estimation, interpolation, and model reduction for applications in signal processing, control theory, and risk assessment.

Jian Li (S'87–M'91–SM'97–F'05) received the M.Sc. and Ph.D. degrees in electrical engineering from Ohio State University, Columbus, in 1987 and 1991, respectively.

From April 1991 to June 1991, she was an Adjunct Assistant Professor with the Department of Electrical Engineering, Ohio State University. From July 1991 to June 1993, she was an Assistant Professor with the Department of Electrical Engineering, University of Kentucky, Lexington. Since August 1993, she has been with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, where she is currently a Professor. Her current research interests include spectral estimation, statistical and array signal processing and their applications.

Dr. Li is a Fellow of IET. She is a member of Sigma Xi and Phi Kappa Phi. She received the 1994 National Science Foundation Young Investigator Award and the 1996 Office of Naval Research Young Investigator Award. She was an Executive Committee Member of the 2002 International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, May 2002. She was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2005, an Associate Editor of the *IEEE Signal Processing Magazine* from 2003 to 2005, and a member of the Editorial Board of *Signal Processing*, a publication of the European Association for Signal Processing (EURASIP), from 2005 to 2007. She has been a member of the Editorial Board of *Digital Signal Processing—A Review Journal*, a publication of Elsevier, since 2006. She is a coauthor of the papers that have received the First and Second Place Best Student Paper Awards at the 2005 and 2007 Annual Asilomar Conference on Signals, Systems, and Computers in Pacific Grove, California. She is a coauthor of the paper that has received the M. Barry Carlton Award for the best paper published in IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS in 2005. She is also a coauthor of the paper that won the Lockheed-Martin Best Student Paper Award at the 2009 SPIE Defense, Security, and Sensing Conference, Orlando, FL, 2009.

Yilu Liu (S'88–M'89–SM'99–F'04) received the B.S. degree from Xian Jiaotong University and the M.S. and Ph.D. degrees from the Ohio State University, Columbus, in 1986 and 1989.

She is currently the Governor's Chair at the University of Tennessee, Knoxville, and Oak Ridge National Laboratory. Prior to joining UTK/ORNL, she was a Professor at Virginia Polytechnic Institute and State University (Virginia Tech). She led the effort to create the North America power grid monitoring network (FNET) at Virginia Tech which is now operated at UTK and ORNL as GridEye. Her current research interests include power system wide-area monitoring and control, large interconnection level dynamic simulations, electromagnetic transient analysis, and power transformer modeling and diagnosis.