Spectrum Combining for ENF Signal Estimation

Adi Hajj-Ahmad, *Student Member, IEEE*, Ravi Garg, *Student Member, IEEE*, and Min Wu, *Fellow, IEEE*

***Abstract*—The Electric Network Frequency (ENF) is the supply frequency of power distribution networks, and is often captured by audio or video measurements recorded near power supplies. The time varying nature of the ENF allows it to be used for such forensic applications as estimating the time and location of media recordings, and discerning their integrity. An initial step in such applications is to extract the ENF signal—instantaneous ENF values over time—as accurately as possible. Existing techniques rely on estimating the ENF around the nominal frequency of 50/60 Hz, or around one of its harmonics at a time. In this letter, a novel spectrum combining approach is proposed, which exploits the presence of the ENF around different harmonics of the nominal frequency. The ENF signal is estimated by combining the ENF at multiple harmonics, based on the local signal-to-noise ratio at each harmonic. A hypothesis testing performance of an ENF-based timestamp verification application is examined to validate that the proposed approach achieves a more robust and accurate performance than conventional ENF estimation techniques.**

***Index Terms*—Electric Network Frequency, timestamp verification, spectrum combining.**

I. Introduction

In recent years, multimedia forensic analysis based on the use of the Electric Network Frequency (ENF) has been demonstrated to be a promising tool for forensic applications. The ENF is the supply frequency in power distribution networks. It is nominally 60 Hz in the United States, and 50 Hz in most other parts of the world. The instantaneous ENF fluctuates about the nominal value due to load control mechanisms within the power grid. We call the changing values of the instantaneous ENF with time an *ENF signal*. These fluctuations can generally be considered a random process.

The ENF signal is used for multimedia forensics applications, as it can be picked up by audio or video recordings made near electrical activity: audio recordings pick up the signal due to mechanical or acoustic hums or electromagnetic interferences from power lines, and video recordings can pick it up due to near-invisible flickering of lightings. It has been shown that ENF-based analysis can be used to estimate the time-of-recording of audio or video files, detect tampering/modification in such files [1]–[4], and determine the location-of-recording among different power grids and potentially within a grid [5], [6].

In most ENF-based analysis cases, it is required to have a database of reference ENF signals. The clean power signal can be measured directly from the power mains through an electrical outlet using a step-down voltage circuit and a recording device. After band-passing the power signal around the nominal frequency, frequency estimation techniques can be applied to extract the frequency estimates on a frame-by-frame basis, thus obtaining the ENF signal. Main types of frequency estimation techniques for ENF extraction include spectrogram-based and subspace-based approaches [5].

The validity of the results obtained in ENF-based studies depends strongly on how well the weak ENF signal is estimated from the available measurements. In power signals, the ENF appears around the nominal frequency of 50/60 Hz and at its harmonics. Most existing frequency estimation approaches for ENF signal extraction rely on the spectral band surrounding the nominal frequency, or on the spectral band surrounding one of the higher harmonics [7], [8].

We have observed that scaled versions of almost the same variations appear in many of the harmonic bands, although the ENF signal strength at different harmonic frequencies differs with recording environments and devices used. An example of this can be seen in Fig. 1 for four settings, which we will explain more in the letter. Following this observation, we propose a low computational complexity approach to extract the ENF signal that strategically makes use of multiple spectral bands. We are inspired by a related problem in handling multipath in wireless communications that has led to a maximum ratio combining approach used in RAKE receivers [9], as well as by a harmonic extension of the classical MUltiple SIgnal Classification (MUSIC) estimator [10]. Our proposed spectrum combining approach exploits different ENF components appearing in a signal, and adaptively combines them based on the local signal-to-noise ratio to achieve a more robust and accurate estimate than that achieved by using only one component. We examine two variants of this approach, based on spectrogram and subspace frequency estimation techniques, respectively. The usefulness of this approach is especially prominent when extracting weak ENF signals from audio/video files, which is challenging due to the presence of noise and media content.

The rest of this letter is organized as follows. Section II explains the model considered and the proposed approach for spectrum combining. Section III discusses the experiments conducted and analyzes the results. Section IV concludes the letter.

II. Model and Proposed Approach

A. Problem Formulation

As motivated in Section I, we devise a technique to estimate the major frequency component of the power signal, through exploiting the presence of the ENF at the base frequency and its harmonics. We estimate, frame-by-frame over time, the deviation Δf_o of the base ENF from the nominal frequency f_o. The same variations of base and harmonic components from the nominal ENF values (up to a scaling factor) provide us multiple

Manuscript received April 24, 2013; revised June 10, 2013; accepted June 19, 2013. Date of publication July 10, 2013; date of current version July 12, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gwenaël Doërr.

The authors are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: adiha@umd.edu; ravig@umd.edu; minwu@umd.edu).

One or more figures in this paper are available in color online at http://http://ieeexplore.ieee.org/

Digital Object Identifier 10.1109/LSP.2013.2272523

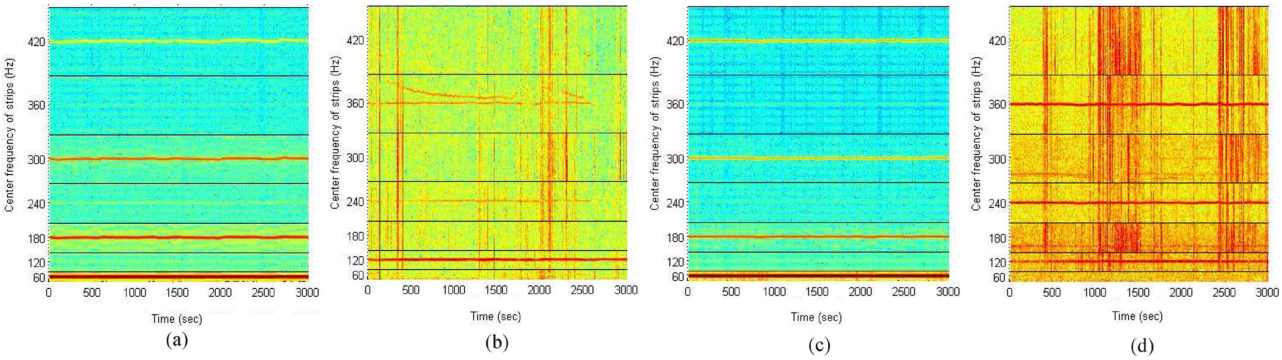

Fig. 1. Spectrogram strips about the harmonics of 60 Hz for two sets of simultaneously recorded power and audio measurements. (a) power recording—residential. (b) audio recording—residential. (c) power recording—workplace. (d) audio recording—workplace.

observations on the base-band deviation Δf_o. To estimate Δf_o for a given time-frame, the estimation can be related to a highly simplified model where we have L observations $Y_1, Y_2, \ldots Y_L$. The kth observation is $Y_k = \alpha_k h(\Delta f_o) + N_k$, where $h(\Delta f_o)$ is a deterministic function of Δf_o representing the spectrum information of the ENF-containing signal that attains its maximum at the ENF component; α_k relates to the strength of the "signal component" around the kth harmonic; the N_k's denote noise components that are assumed to be independently distributed, each following a normal distribution with mean zero and variance σ_k^2. To estimate the function values $h(\Delta f_o)$, a Maximum Likelihood Estimator (MLE) can be adopted as:

$$\hat{h}_{MLE}(\Delta f_o) = \frac{\sum_k (Y_k/\alpha_k) \cdot (\alpha_k^2/\sigma_k^2)}{\sum_k (\alpha_k^2/\sigma_k^2)}. \quad (1)$$

This suggests producing an ENF estimate by combining multiple base and harmonic spectral bands, each weighted according to its relative strength with respect to noise; the combined spectrum can then be processed to obtain Δf_o.

In the spectral domain, the ENF-containing signal can be considered a summation of impulses at the base ENF and its harmonics. Due to the recording mechanism and environment, additional frequency components may interfere around these bands, and we consider such interferences as noise around each harmonic. For a given time-frame, the observed power spectrum component, $P_{B,k}(f)$, contributed at the kth harmonic band, analogous to Y_k in (1), can be expressed as

$$P_{B,k}(f) = A_k h_k(f) + P_{n,k}(f), \quad (2)$$

for $f \in [k(f_o - f_B), k(f_o + f_B)]$, where f_B reflects the empirical support of ENF presence around the base frequency. Here, A_k denotes the magnitude of the energy contributed by the frequency component close to kf_o; $P_{n,k}(f)$ denotes the independent noise component around the kth harmonic, assumed white within the bandwidth of interest; $h_k(f)$ denotes an impulse-like function that attains its maximum at $f = k(f_o + \Delta f_o)$. In practice, we observe $h_k(f)$ as a peak energy concentration with a small spread. Assuming that the power signal contains L harmonics, we can write the power spectrum of the signal for a given time-frame as $P_{signal}(f) = \sum_{k=1}^{L} P_{B,k}(f)$.

To estimate the frequency deviation Δf_o for a given time-frame, the proposed spectrum combining approach first compresses and shifts the spectrum components to the nominal base range $[f_o - f_B, f_o + f_B]$, and then combines the components together. This is analogous to (1), where the weighted summation can then be written as:

$$S(f) = \sum_{k=1}^{L} w_k P_{B,k}(kf). \quad (3)$$

When applying (3), the $h_k(f)$ components of (2) are compressed along the frequency axis to become $h_k(kf)$ components; they should each have their maximum at $f = f_o + \Delta f_o$. The combining weight, w_k, in (3) has been introduced to weigh the various harmonic spectral bands based on the signal-to-noise ratio (SNR) around the corresponding harmonic. This weight is analogous to the α_k^2/σ_k^2 in the numerator of (1); the denominator in (1) can be seen as a normalization parameter for these SNR-based combining weights. The frequency $f_{ENF} = f_o + \Delta f_o$ can be obtained by searching for the maximum in $S(f)$.

B. Determining Spectral Combining Weights

The combining weights are computed for a recording over a certain time duration (e.g., 30 minutes), and then recomputed for subsequent durations. This makes the weights adaptive to reflect the changing strength of the ENF in the bands over time. Taking the combining weight w_1 for the base band around f_o for a certain time duration as an example, we set w_1 as an estimate of the SNR in the band; $\hat{P}_{signal,1}/\hat{P}_{noise,1}$. We choose $f_B = 1$ Hz, because in the U.S., the ENF fluctuates within 59.98 Hz and 60.02 Hz. The overall band for computing SNR would be [59, 61] Hz. Then, $\hat{P}_{signal,1}$ is the average power spectral density (PSD) within the band [59.98, 60.02] Hz over the chosen time duration and $\hat{P}_{noise,1}$ is the average PSD in the bands [59, 59.98] Hz and [60.02, 61] Hz over the same time duration.

The combining weights for other harmonic spectral bands are computed in the same manner, except for different sizes of the spectral bands considered.

C. Instantaneous ENF Estimation

After the w_k's are computed, the spectrum $S(f)$ can be computed using (3). $S(f)$ can be seen as a weighted summation of shifted and compressed spectral bands. We explore the estimation of spectral bands through two methods, a spectrogram and a subspace-based "pseudo-spectrum". In the first method, the spectral harmonic bands are chosen from the spectrogram and shifted and compressed to compute $S(f)$. In the second method, a subspace-based frequency estimation technique such as MUSIC is used. The MUSIC algorithm looks for P complex exponential frequencies in a signal, and computes a pseudo-spectrum [11]. The peaks in the pseudo-spectrum correspond to the frequencies found in the signal. Here, we apply band-pass filters on the ENF-containing signal around the ENF harmonics. For each band-passed signal, the pseudo-spectrum can be computed for $P = 2$ (corresponding to the positive and negative ENF), and the spectral band around the harmonic of interest is identified from the computed pseudo-spectrum and stored. After all harmonic bands are estimated, they are used to compute $S(f)$.

In practice, when computing $S(f)$, we shift the spectral bands to one of the higher bands rather than the base band, compressing or expanding as necessary, to make use of the wider range of variations available around higher harmonics for the same frequency resolution. In the results shown in Section III, we shift the bands to the 240 Hz band. After $S(f)$ is computed, we search for its maximum through quadratic interpolation [12]. The ENF is set as the argument of the maximum. If $S(f)$ is defined around a higher spectral band than the base band, the solution is scaled accordingly.

III. Experiments and Results

A. Experimental Set-Up

We carry out recordings in two different environment settings, one overnight in a residential setting (an apartment) and one during the day in a workplace setting (an office). Each set of recordings is composed of a power mains signal, used as a reference signal, and an audio signal recorded concurrently and expected to pick up the ENF. The audio is recorded using a battery powered Olympus Voice Recorder WS-700 M at a sampling rate of 44.1 kHz in MP3 format at 256 kbps. All recordings are made in Maryland, which is part of the US Eastern Grid. The recordings are downsampled to 1000 Hz in WAV format to ease the computations. Their spectra are estimated for consecutive frames of 5 seconds long each. Sample spectrogram strips around the harmonics of the nominal ENF for each of the four recordings are shown in Fig. 1.

From Fig. 1(a) and (c), we can see that the power signal is almost noise free around the harmonics, and the ENF has a strong presence around the odd harmonics. Fig. 1(b) shows that in the residential audio recording, the ENF is present strongly only around 120 Hz, with faint to no presence around the other harmonics. On the other hand, Fig. 1(d) shows stronger presence for the ENF in the workplace audio signal around 240 Hz and 360 Hz; the noisy component appearing around 120 Hz is not the ENF as will be shown in Section III-C. Interestingly, the ENF has almost no presence around the nominal 60 Hz in the audio signals, which can be due to either the recording environments or the recorder used or both. This shows the varied presence and strength of the ENF at harmonics in recordings made under different conditions.

We estimate the ENF from the four recordings using our proposed spectrum combining approach. Table I shows sample values of combining weights computed at various harmonics considered for the four recordings. The combining weights shown are normalized, each expressed as a percentage of the sum of combining weights of all the spectral bands for its time duration. The resulting values conform with our earlier observations on the bands where the ENF has a strong presence. We also noticed that in some cases, the weight before normalization is less than 1, which implies that in such bands, the noise component is stronger than the signal component. We set such weights to zero before normalization and thus exclude them from the summation of (3).

TABLE I
Sample Values of Spectral Combining Weights w_k's

Center Freq. (Hz)	Power Residential	Power Office	Audio Residential	Audio office
60	6.32	6.81	0	0
120	2.03	1.02	75.75	15.69
180	18.79	19.75	0	1.80
240	4.73	1.99	13.77	32.00
300	28.88	29.15	5.20	1.91
360	2.70	4.07	5.28	44.69
420	35.61	36.16	0	2.19
480	0.93	1.05	0	1.72

B. Comparison Framework

To assess the performance of our proposed spectrum combining approach, we compare it to the conventional approach of using frequency estimation techniques on a single spectral band. Since our audio recordings show a strong presence of the ENF at 120 Hz, 240 Hz and/or 360 Hz, we generate ENF estimates based on the individual bands centered around these frequencies. We observe that isolated outliers may appear in the ENF estimates outside the known range of ENF variations ([59.98, 60.02] Hz for the nominal band); we replace these outliers by the average of the frequency estimates preceding and succeeding them. For each audio ENF signal estimate, we have a corresponding ENF signal estimate extracted from a power signal recorded simultaneously with the audio signal. We generate the audio ENF estimates twice, once using the spectrogram-based technique and once using MUSIC. For the reference power ENF, we use the average of the estimates obtained through the two techniques. As mentioned earlier, in applying both techniques on the signals, we find the maximum in the spectrogram or pseudo-spectrum through quadratic interpolation.

To compare between ENF estimates from an audio signal and a reference power signal, we split both ENF signal estimates into segments of 96 points each, corresponding to 8 minutes of data. We examine the performance in ENF estimation for time-of-recording authentication, a main application of ENF [2], [3]. We consider a hypothesis testing framework:

$$\begin{cases} H_0: & \text{segments were recorded at different times.} \\ H_1: & \text{segments were recorded simultaneously.} \end{cases}$$

We find the correlation coefficient between all combinations of segments from both ENF signals, and apply thresholding to decide on H_0 vs. H_1. Figs. 2 and 3 show the Receiver Operator Characteristic (ROC) curves for the cases studied.

C. Results and Discussions

Figs. 2 and 3 show that the spectrum combining approach outperforms the approaches for individual bands; it behaves comparably to estimation around the "dominant harmonic" if present, i.e., 120 Hz in the residential recordings and 360 Hz in the workplace recordings. Figs. 2(b) and (b) demonstrate the ability of the proposed approach to suppress spurious peaks appearing at certain harmonics due to noise and distortions: the poor ROC curve from estimation around 120 Hz reveals that the frequency component in that band, seen in Fig. 1(d), is not

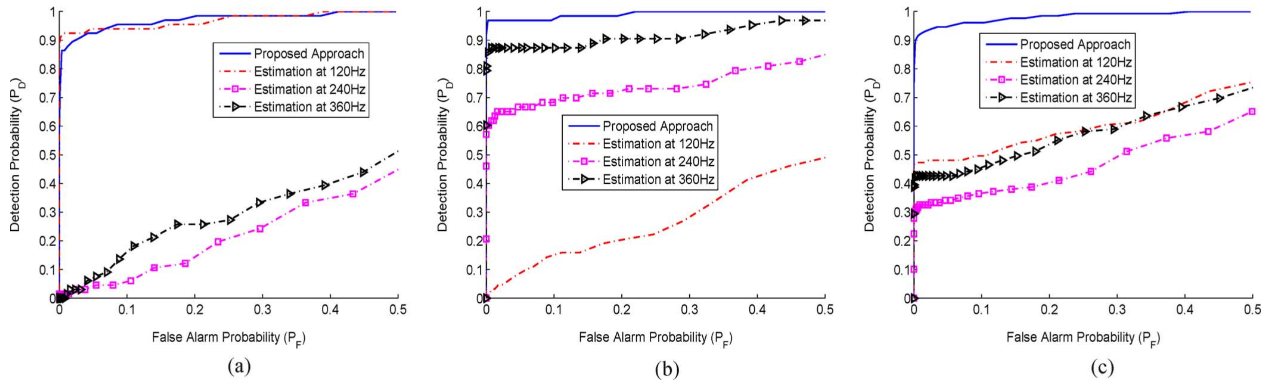

Fig. 2. Spectrogram-based results: ROC curves for matching ENF estimates from audio signals to those estimated from reference power signals. (a) ROC for residential data. (b) ROC for workplace data. (c) ROC for both sets of data together.

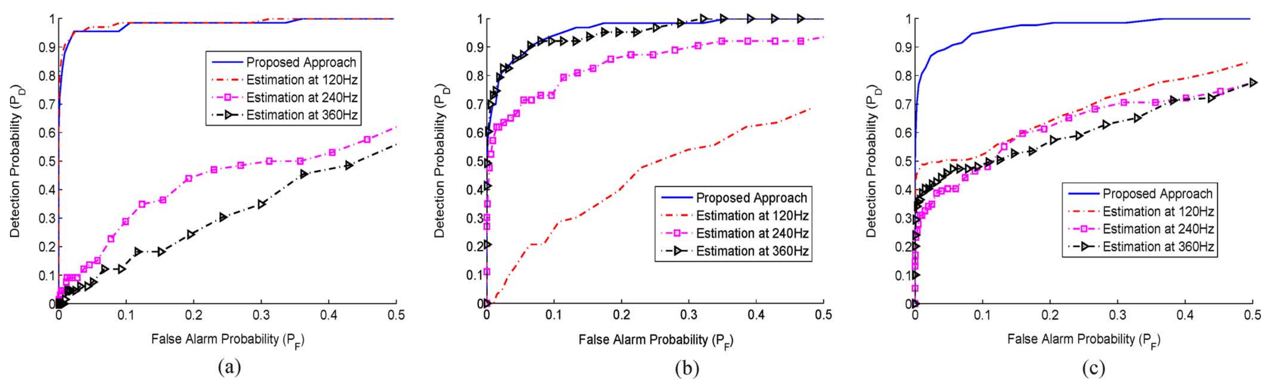

Fig. 3. MUSIC-based results: ROC curves for matching ENF estimates from audio signals to those estimated from reference power signals. (a) ROC for residential data. (b) ROC for workplace data. (c) ROC for both sets of data together.

the ENF. It could be the result of stray electromagnetic fields with complex spectra found in the workplace (office) setting [7]. The proposed approach was unaffected and was able to leverage the true ENF components at 240 Hz and 360 Hz to achieve a good ENF estimate. Figs. 2(c) and 3(c) demonstrate the proposed approach's robustness and its ability to adapt to unpredictable changes in ENF strengths at harmonics in cases where the dominant harmonic varies across recordings.

Comparing the two variants of the proposed approach, we can see that the spectrogram-based method yields better results than the MUSIC-based method. As the MUSIC pseudo-spectrum is computed through a rather sophisticated procedure and its physical meaning is not as direct as the spectrogram's, the observations from MUSIC may thus deviate more from the model discussed in Section II-A than the spectrogram-based observations do. This would affect the effectiveness of the spectrum combining, and explain the discrepancy in the performance between the two variants.

IV. Conclusions and Future Work

In this letter, we propose a novel spectrum combining approach for extracting the ENF signal from multimedia recordings. The approach makes use of the ENF around the different harmonics of 50/60 Hz rather than only one harmonic. This is achieved through a weighted summation of multiple spectral bands from around the harmonics, weighted according to the local SNR in each band. Our experiments have shown that the proposed approach can achieve more accurate and robust estimates than the conventional approach for ENF estimation. Of the two variants presented, the improvement in performance was more visible in the spectrogram-based variant over the MUSIC-based variant. In our future work, we plan to study this performance difference in more detail.

References

[1] C. Grigoras, "Applications of ENF analysis in forensic authentication of digital and video recordings," *J. Audio Eng. Soc.*, vol. 57, no. 9, pp. 643–661, Sep. 2009.

[2] R. Garg, A. L. Varna, and M. Wu, "Seeing ENF: natural time stamp for digital video via optical sensing and signal processing," in *Proc. 19th ACM Int. Conf. Multimedia*, Scottsdale, AZ, 2011, pp. 23–32.

[3] R. Garg, A. L. Varna, A. Hajj-Ahmad, and M. Wu, "Seeing ENF: Power signature based timestamp for digital multimedia via optical sensing and signal processing," *IEEE Trans. Inf. Forensics Secur.*, 2013, 10.1109/TIFS.2013.2272217 , to be published.

[4] D. Rodriguez, J. Apolinario, and L. Biscainho, "Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 3, pp. 534–543, Sep. 2010.

[5] A. Hajj-Ahmad, R. Garg, and M. Wu, "Instantaneous frequency estimation and localization for ENF signals," in *Proc. Asia Pacific Signal & Info. Processsing Association*, Los Angeles, CA, USA, 2012.

[6] R. Garg, A. Hajj-Ahmad, and M. Wu, "Geo-location estimation from electric network frequency signals," in *Proc. Int. Conf. Acoustics, Speech & Signal Processing*, Vancouver, BC, Canada, 2013.

[7] E. B. Brixen, "Further investigation into the ENF criterion for forensic authentication," *Audio Eng. Soc. Conv. 123*, 2007.

[8] O. Ojowu, J. Karlsson, J. Li, and Y. Liu, "ENF extraction from digital recordings using adaptive techniques and frequency tracking," *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 4, pp. 1330–1338, Aug. 2012.

[9] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[10] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Subspace-based fundamental frequency estimation," in *Proc. Eur. Signal Processing Conf.*, Vienna, Austria, 2004, pp. 637–640.

[11] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York, NY, USA: Wiley, 1996.

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