IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 December 2023, accepted 10 January 2024, date of publication 15 January 2024, date of current version 22 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3354053

RESEARCH ARTICLE

Leveraging Electric Network Frequency Estimation for Audio Authentication

CHRISTOS KORGIALAS[1], (Graduate Student Member, IEEE), CONSTANTINE KOTROPOULOS[1], (Senior Member, IEEE), AND KONSTANTINOS N. PLATANIOTIS[2], (Fellow, IEEE)

[1]Department of Informatics, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece
[2]Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada

Corresponding author: Christos Korgialas (ckorgial@csd.auth.gr)

This work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "2nd Call for H.F.R.I. Research Projects to support Faculty Members and Researchers" under Project 3888.

ABSTRACT The Electric Network Frequency (ENF) serves as a simple means to verify the authenticity of audio recordings. ENF variations contain crucial information, acting as a distinctive "fingerprint" when electronic devices are connected or located near power mains. A novel framework for ENF estimation is proposed. This approach alternates between the Least Absolute Deviation (LAD) regression for determining regression weights and objective function minimization with respect to frequency, adapting them within the context of the ℓ_1 norm or the sum of ℓ_1 norms of the approximation error. This framework is a direct consequence of Laplacian distributed noise. Goodness-of-fit tests are reported, indicating that the Laplacian noise hypothesis is more appropriate than the hypothesis of Gaussian noise in the benchmark ENF-WHU dataset. Extensive evaluation using audio recordings from the aforementioned dataset demonstrates the exceptional performance of the proposed framework outperforming state-of-the-art ENF estimation schemes. These findings provide compelling evidence for the efficacy of the proposed ENF estimation schemes as reliable prerequisites for detecting audio forgeries.

INDEX TERMS Electric network frequency (ENF), robust ENF estimation, least absolute deviations (LAD) regression, audio authentication, multimedia forensics.

I. INTRODUCTION

With the rapid advancements in speech synthesis and voice conversion technologies, the landscape of audio recordings has undergone a significant transformation. Fake audio recordings that closely resemble human speech can now be easily generated, highlighting the need for rapid and precise tools to identify them. What adds to the complexity of this issue is the emergence of Artificial Intelligence (AI)-powered text-to-audio machines, which further amplify the challenge of distinguishing genuine recordings from fraudulent ones.

In this context, leveraging Electric Network Frequency (ENF) estimation constitutes a prerequisite for legitimate audio verification. ENF serves as a unique environmental fingerprint embedded within audio recordings that are captured near the power mains [1]. By accurately extracting and analyzing the ENF variations present in the recordings, it is possible to discriminate genuine audio content from manipulated or fabricated ones.

ENF fluctuates instantaneously around its nominal value of 60 Hz in the United States/Canada or 50 Hz in other parts of the world. At any time instant, the ENF exhibits almost the same fluctuation across an interconnected power network. Thus, the ENF signal acquired from any power outlet in such a network during a particular time period can be utilized as a reference signal (i.e., ground truth) to be attested with the ENF extracted from the multimedia recordings [2], [3]. ENF gets intrinsically integrated into audio recordings by a dynamic microphone near mains-powered devices or to transmission cables due to electromagnetic wave

The associate editor coordinating the review of this manuscript and approving it for publication was Deepak Mishra.

propagation [4]. The presence of acoustic mains hum, which is generated by various equipment, including typical household appliances, causes the ENF to become embedded in an audio signal recorded by an electret microphone [5], [6].

While numerous methods [7], [8] have been developed to detect audio fakes, ENF extraction stands out as a distinctive and powerful forensic tool in this domain. The use of ENF as an authentication tool [9], [10] has proven to be highly effective in verifying the authenticity of multimedia recordings, offering a robust method to combat deep fake attacks and ensure the integrity of digital content. ENF has been exploited in multimedia forensics and anti-forensics analysis [11], enabling timestamp verification [12], [13], [14] and geo-location estimation [15], [16].

Despite the extensive research efforts in the ENF-based multimedia forensics field [17], several challenges persist [18]. These include the need for an accurate estimation of the ENF in short audio recordings. Another challenge is the extraction of ENF, which is much weaker than the noise and audio content in a recording. The ENF extraction involves estimating the instantaneous frequency (IF) by segmenting the recording into overlapping frames. Interference and low signal-to-noise ratio (SNR) hinder ENF estimation. Additionally, normal device movement can introduce significant Doppler effects and eliminate the ENF signal in audio [19]. To address these limitations, a notable approach is the Graph-based Harmonic Selection Algorithm (GHSA) [20], which finds the optimal combination of harmonic components for ENF estimation. The Harmonic Robust Filtering Algorithm (HRFA) is also utilized to extract the ENF from noisy observations. The Maximum Likelihood ENF estimators (MLE) [21] and Weighted MLE (WMLE) [22], incorporating both the GHSA and HRFA algorithms, are referred to as P-MLE and P-WMLE, respectively. These ENF estimators along with others in [20] constitute the state-of-the-art methods for the problem addressed in the paper.

Here, assuming Laplacian noise, ENF estimation is leveraged from the perspective of Least Absolute Deviation (LAD) regression to find the regression weights and objective function minimization with respect to frequency adapting the ENF estimation schemes in [20], [21], and [22] in the context of the ℓ_1 norm or the sum of ℓ_1 norms of the approximation error. To enhance the ENF estimation accuracy, 10 novel single/multi-tone ENF estimation schemes are developed (see Section IV). A fair comparison with the experiments conducted in [20] demonstrates an improved accuracy in ENF estimation. The major contribution of the paper lies in the formulation of ENF estimation as a LAD regression, which alternates between the solution for the regression weights and the minimization with respect to the frequency of objective functions resorting to the ℓ_1 norm or the sum of ℓ_1 norms of the approximation error until convergence. The performance of the proposed ENF estimation schemes is evaluated on the ENF-WHU dataset [20], consisting of 130 audio recordings. The validity of the Laplacian noise hypothesis is thoroughly assessed with Goodness-of-Fit tests at the noise model that emerged from the signal model and the regression approximation error. Objective figures of merit based on the Mean Square Error (MSE) are employed, such as the average MSE (AMSE) and the standard deviation of MSE across all frames of a recording. The authentication of audio recordings is assessed by calculating and reporting the AMSE between the estimated ENF extracted from the audio recordings and the ground truth ENF. To our knowledge, this is the first time ENF estimation is treated as a LAD regression problem, extending the use of LAD regression for ENF detection [23].

TABLE 1. List of abbreviations.

Abbreviation	Expansion
AI	Artificial Intelligence
AMSE	Average Mean Square Error
AMTC	Adaptive Multi-Trace Carving
CC	Correlation Coefficients
ENF	Electric Network Frequency
FFT	Fast Fourier Transform
GHSA	Graph-based Harmonic Selection Algorithm
HRFA	Harmonic Robust Filtering Algorithm
IF	Instantaneous Frequency
LAD	Least Absolute Deviation
LASSO	Least Absolute Shrinkage and Selection Operator
LS	Least Squares
MLE	Maximum Likelihood Estimator
MSE	Mean Square Error
RFA	Robust Filtering Algorithm
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier Transform
WMLE	Weighted Maximum Likelihood Estimator

The novel contributions of the paper are as follows:

1) The ENF estimation is addressed from the perspective of LAD regression.
2) Ten novel ENF estimation schemes are developed by adapting those in [20], [21], and [22] in the context of minimization with respect to frequency of objective functions employing the ℓ_1 norm or sum of ℓ_1 norms of the approximation error.
3) Goodness-of-Fit tests performed on the ENF-WHU dataset at both the noise model and the regression approximation error validate the hypothesis of Laplacian noise.
4) A thorough experimental assessment of the ENF-WHU dataset demonstrates that the proposed ENF estimation schemes outperform the state-of-the-art ENF estimation schemes in [20].

The remainder of the paper is organized as follows. In Section II, prior work is surveyed. The ENF fundamentals are detailed in Section III. The problem formulation and the proposed framework are presented in Section IV. The performance of the proposed framework is evaluated in

Section V. In Section VI, conclusions are drawn, and future work is suggested. To enhance readability, a list of abbreviations is provided in Table 1.

II. PRIOR WORK

A. ENF FOR AUDIO AUTHENTICATION

The incorporation of ENF serves as an exceptional and dependable feature that effectively verifies the integrity and source of audio recordings, significantly bolstering the authenticity and instilling trust. In [24], the authenticity of audio recordings was established by comparing the logged variations of the ENF in the mains hum of the questioned recording to reference ENF values. An audio authenticity detection algorithm, based on the Max Offset for Cross-Correlation between the extracted ENF signal and the reference signal, was proposed in [25]. The ENF signal was extracted from a query audio signal and partitioned into overlapping blocks for forgery detection. In [26], an ENF-based audio authentication system examined audio recordings by matching the ENF signal from a questioned recording with the reference signal stored in a database for timestamp verification. Also, the Absolute Error Map was introduced to detect tampering and explore audio authentication. In [27], the ENF criterion was employed to authenticate digital audio recordings in legal proceedings using a wide-area Frequency Monitoring Network as the reference frequency database and a modified Short-Time Fourier Transform (STFT) for ENF estimation. A forensic tool was proposed in [28] to assess audio authenticity by detecting phase discontinuity of the power grid signal (i.e., the ENF). This tool was utilized to estimate editing points and make automatic decisions regarding the authenticity of the audio evidence.

B. ENF ESTIMATION

ENF extraction in audio recordings has garnered significant research interest over the years, positioning it as a prominent forensic tool. In [21], a multi-tone harmonic model for the ENF estimation was described. To estimate the ENF signal more accurately, many harmonics were merged, and the Cramer-Rao bound was applied to limit the variance of the ENF estimator. A spectral estimation approach that combined the ENF at multiple harmonics was discussed in [22]. The ENF was extracted considering the local SNR at each harmonic. The ENF extraction was treated as a data-dependent (adaptive) filtering problem instead of the conventional STFT [29]. This method resulted in high-resolution results but at a significant computational cost. An ENF extraction algorithm of training audio and power recordings from different grids was proposed in [30]. The STFT was employed to correct the erroneously selected ENF peaks by leveraging time correlations. In [31], the Adaptive Multi-Trace Carving (AMTC) approach was developed to detect and track subtle frequency components in noisy signals through iterative dynamic programming and adaptive trace compensation. In [32], a binary approach was introduced to desired specific spectral lines instead of the entire frequency band. The main core of ENF extraction was the discrete Fourier transform algorithm. In the [33], a non-parametric approach for ENF estimation was developed, which incorporated a custom lag window design into the Blackman-Tukey spectral estimation method. In [34], a Hilbert-based transform for instantaneous frequencies estimation was described for estimating the instantaneous frequencies. The insights from [35] and [36] were exploited to develop a Capon spectral estimation applied to ENF estimation leveraging the Gohberg-Semencul factorization [37]. Furthermore, incorporating a Parzen temporal window emphasized the significance of window selection for accurate estimation of ENF. The combined methodology not only achieved lower computational complexity but also improved spectral resolution, leading to more accurate ENF estimation. In [38], the ENF extraction was addressed as a frequency demodulation problem. The ENF was approached as a sinusoid at the nominal frequency by creating and analyzing an IF signal instead of the direct measurements. The research conducted in [39] was focused on estimating the frequencies in short time frames. A systematic examination of a number of high-resolution, low-complexity frequency estimation techniques was employed. A decorrelation algorithm for ENF estimation from a recaptured audio was proposed in [40]. The dominant values of ENF were subtracted from the original signal, resulting in a latent ENF extraction. In [41], special emphasis was given to the preprocessing stage. Principal component analysis was applied to eliminate the interference from speech content prior to ENF estimation. The method successfully combined the fundamental ENF and its harmonics for very short audio clips. In [42], a filtering method termed Robust Filtering Algorithm (RFA) was introduced to deal with noise interference. To obtain a time-frequency representation suitable for ENF estimation, a kernel function was used.

C. ENF DETECTION

Although significant attention has been paid to ENF estimation in audio recordings, a topic that has not been addressed adequately is ENF detection. That is, whether ENF is present or not in a recording. A notable exception was the detailed study in [43]. That study introduced six different detectors, of which three were evaluated in real-world audio recordings. The detectors showed a credible performance for short audio clips, enabling reliable ENF detection. In [44], a multi-tone time-frequency detector was developed by utilizing a multi-harmonic combination to determine if valid ENF traces were present in a recording as well as to offer information on the overall ENF quality and the number of accessible harmonic components. In [23], a LAD-Likelihood Ratio Test ENF detector was proposed that improved the accuracy and robustness of ENF detection in short-length recordings compared to the state-of-the-art detectors.

In [45], a superpixel-based ENF presence detector for digital video was developed. Several ENF signal estimates

from stable superpixel areas identified whether an ENF signal was present or absent in short video clips. An automated ENF disturbance detector using a linear discriminant was proposed in [46]. ENF extraction was performed using the Estimation of Signal Parameters by Rotational Invariant Techniques before evaluating the detector.

III. ENF FUNDAMENTALS

Catalin Grigoras introduced the ENF criterion in forensic analysis [47]. Differences in power production and consumption cause the ENF variations. In Europe, at any given time, the ENF can be expressed as $f = 50 \pm \Delta f$ Hz, where Δf signifies the aforementioned ENF fluctuations around the nominal frequency [4]. Considering Δf, there are three types of network operating conditions [48]. If $\Delta f \leq 50$ mHz, the conditions are considered to be normal. If Δf is between 50 and 150 mHz the operating conditions are deemed to be impaired, but with no major risk. If $\Delta f \geq 150$ mHz, the operating conditions are deemed to be severely impaired, resulting in significant risks of malfunction of the electric network.

Following [4], a threefold ENF extraction approach can be pursued.

- Time and frequency domain: The spectrogram of short-time recordings is derived and compared visually with the reference ENF signals.
- Time domain: After proper Finite Impulse Response filtering, zero-crossings are measured around the frequency of interest.
- Frequency domain: The periodogram of short-time segments and their spectral peaks (i.e., searching for magnitude peak around 50 Hz) are computed through the Fast Fourier Transform (FFT). Each spectral peak is compared against the ground truth ENF signal.

The spectrogram approach is the quickest and easiest to employ [4]. It reveals the ENF components (i.e., harmonics) and is particularly helpful for comparing the questioned ENF dates and times against the database of ENF dates and times. Only one ENF component can be extracted in the time domain approach. The frequency domain approach estimates one ENF component as well. The proposed framework, detailed in Section IV, is a frequency domain approach.

IV. PROBLEM FORMULATION AND METHODOLOGY

In this Section, starting from signal modeling, ENF estimation schemes are derived. In Section IV-A the audio feed is defined. A brief description of the state-of-the-art ENF estimation schemes can be found in Section IV-B. The proposed schemes are analyzed in Section IV-C.

A. AUDIO FEED

The initial step is to define the audio feed after bandpass filtering that retains the signal content around the ENF harmonics. Following [20], the filtered audio signal $x[n]$, $n \in \{0, 1, \ldots, N-1\}$, is approximated as

$$\begin{aligned} x[n] &= s[n] + v[n] \\ &= \sum_{m \in \mathcal{M}} A_m[n] \cos\left(2\pi T m f[n] + \phi_m\right) + v[n], \end{aligned} \tag{1}$$

where $s[n]$ represents the ENF signal, while $v[n]$ denotes colored Gaussian noise as a consequence of the bandpass-filtered white Gaussian noise. Later on, the assumption of Gaussian distributed noise will be challenged. In (1), $A_m[n] > 0$ and ϕ_m is the time-varying amplitude and phase of the m-th harmonic, $mf[n]$, with $f[n]$ denoting the IF (i.e., the nominal ENF frequency), respectively. The term $T = 1/f_S$ refers to the sampling interval with f_S being the sampling frequency. In (1), the ENF is treated as an unknown deterministic signal that follows the multi-tone harmonic model consisting of $\mathcal{M} = |\mathcal{M}|$ harmonic components with $\mathcal{M} \subseteq \mathbb{Z}_+$ being the set of harmonic indices.

The direct IF calculation from (1) is a challenging task because of the signals' time-varying nature. This limitation is addressed thanks to the STFT, which allows for analyzing the signal's frequency content over short overlapping frames, providing a time-varying representation of the signal's spectrum. As a result, each IF is treated as a piecewise constant for each frame. Let N_{ENF} denote the number of frames that emerged, and N_F be the length of each frame. If Δ denotes the number of samples each frame is advanced (i.e., the frame step-size), N_{ENF} is given by

$$N_F + (N_{\text{ENF}} - 1)\,\Delta = N \Rightarrow N_{\text{ENF}} = \frac{N - N_F}{\Delta} + 1, \tag{2}$$

where N is the length of $x[n]$. N_{ENF} is the length of ENF signal $f[l]$ (i.e., the IF time series). The l-th frame, $l \in \{0, 1, \ldots, N_{\text{ENF}} - 1\}$ is expressed as [20]:

$$\begin{aligned} x_l[n] &= s_l[n] + v_l[n] \\ &= \sum_{m \in \mathcal{M}} A_{m,l}[n] \cos\left(2\pi T m f[l] n + \phi_{m,l}\right) + v_l[n], \end{aligned} \tag{3}$$

where $n \in \{0, 1, \ldots, N_F - 1\}$ and $f[l]$ is the constant frequency for the l-th frame.

B. INSTANTANEOUS FREQUENCY ESTIMATION

After determining the audio feed $x_l[n]$ in (3), an additional step prior to ENF estimation is harmonic enhancement. This is achieved by employing either the RFA [42] or the HRFA [20] for single-tone or multi-tone enhancement, respectively. These algorithms enhance the harmonic components present in the ENF signal.

Let $\hat{f}_{\text{single}}[l] = \text{argmax}_f\, P_l(f)$ denote the frequency corresponding to the spectral peak of the periodogram of $x_l[n]$[1] given by

$$P_l(f) = \left| \sum_{n=0}^{N_F - 1} x_l[n] e^{-j 2\pi T f n} \right|^2, \quad |f| \leq \frac{f_S}{2}. \tag{4}$$

[1]Note that the constant factor $1/N_F$ prior to the squared magnitude is omitted in (4) without harming the analysis.

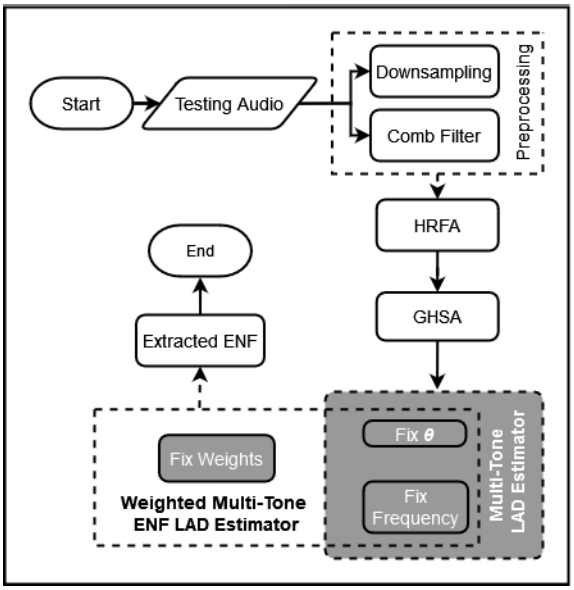


FIGURE 1. Flowchart of the proposed LAD-based ENF estimation framework.

If the RFA [42] is applied to $x_l[n]$, then its output is referred to as $\breve{x}_{E,l}[n]$. Searching for the frequency corresponding to the spectral peak of the periodogram of $\breve{x}_{E,l}[n]$ yields $\hat{f}_{\mathrm{E-single}}[l]$.

Let $\hat{f}_{\mathrm{MLE}}[l]$ be the outcome of the search within the periodogram for the maximum sum-of-squares [21], i.e.,

$$\hat{f}_{\mathrm{MLE}}[l] = \underset{f}{\operatorname{argmax}} \sum_{m\in\mathcal{M}} P_l(mf). \tag{5}$$

If $w_{m,l}$ denote the SNR of the m-th harmonic in the l-th frame, WMLE searches for the maximum sum-of-weighted-squares [22]:

$$\hat{f}_{\mathrm{WMLE}}[l] = \underset{f}{\operatorname{argmax}} \sum_{m\in\mathcal{M}} w_{m,l}\, P_l(mf). \tag{6}$$

Let $x_{\mathrm{E},l}[n]$ denote the output of HRFA [20], $x_{\mathrm{S},l}[n]$ be the output of GHSA [20], and $x_{\mathrm{P},l}[n]$ refer to the output of sequential application of HRFA and GHSA. By adapting (5) or (6) in the context of $x_{\mathrm{E},l}[n]$, $x_{\mathrm{S},l}[n]$, $x_{\mathrm{P},l}[n]$, six ENF estimation schemes result, namely $\hat{f}_{\mathrm{E-MLE}}[l]$, $\hat{f}_{\mathrm{E-WMLE}}[l]$, $\hat{f}_{\mathrm{S-MLE}}[l]$, $\hat{f}_{\mathrm{S-WMLE}}[l]$, $\hat{f}_{\mathrm{P-MLE}}[l]$, and $\hat{f}_{\mathrm{P-WMLE}}[l]$.

Let the periodogram of the $x_{\mathrm{E},l}[n]$ be

$$P_{\mathrm{E},l}(f) = \left| \sum_{n=0}^{N_{\mathrm{F}}-1} x_{\mathrm{E},l}[n] e^{-\jmath 2\pi T f n} \right|^2. \tag{7}$$

The goal of P-MLE is to find

$$\hat{f}_{\mathrm{P-MLE}}[l] = \underset{f}{\operatorname{argmax}} \sum_{m\in\Omega} P_{\mathrm{E},l}(mf), \tag{8}$$

where $\Omega \subseteq \mathcal{M}$ is the subset of harmonic components with the highest mutual correlation coefficients (CC) selected by the GHSA [20]. The selection is treated as a maximum weight clique problem in graph theory, using the Bron-Kerbosch algorithm [49]. It is seen that (8) is a maximum sum of squares optimization problem where all the harmonic components are taken into account. Alternatively, one may resort to a single-tone Ω (i.e., $|\Omega| = 1$). Let $\{f[l]\}_m$ be the IFs in each harmonic frequency band where $\{\hat{\cdot}\}$ is omitted for notation simplicity. The smoothest harmonic component is obtained by

$$m^* = \underset{m\in\mathcal{M}}{\operatorname{argmin}} \sum_{l=1}^{N_{\mathrm{ENF}}-1} |\{f[l]\}_m - \{f[l-1]\}_m| \tag{9}$$

and used in (8).

P-WMLE searches for the maximum sum-of-weighted-squares. In this method, the weights represent the relative energy in the signal subband compared to the energy in the noise subband, where higher weights indicate a higher local SNR and a greater contribution to the ENF estimation [20], i.e.,

$$\hat{f}_{\mathrm{P-WMLE}}[l] = \underset{f}{\operatorname{argmax}} \sum_{m\in\Omega} w_{m,l}\, P_{\mathrm{E},l}(mf), \tag{10}$$

where $w_{m,l}$ is the weight for the m-th harmonic in the l-th frame of the enhanced signal $x_{\mathrm{E},l}[n]$.

In the following, the ENF estimate for each frame using the scheme $\{\sim\}$ is denoted as $\hat{f}_{\{\sim\}}[l]$, where l stands for the frame index of the enhanced signal $x_{\mathrm{E},l}[n]$. There are 10 schemes in total.

C. PROPOSED FRAMEWORK

Here, the description of the proposed framework for ENF estimation shown in Figure 1 is detailed. Initially, the audio feed is preprocessed to retain the harmonics of ENF, which are subsequently enhanced using the HRFA [20]. Then, the harmonics with stronger ENF components are selected through the GHSA [20]. The estimation process is alternately optimizing for the LAD coefficients and searching for $\hat{f}_{\{\sim\}}[l]$ by fine grid searching.

To begin with, the proposed framework resorts to the assumption that the noise after centering in (3) follows a Laplacian distribution Laplace$(0, b)$, where b is a scale parameter. To determine which dictates the distribution's spread, a data-driven approach is employed. The same applies to the mean of the noise samples. The aforementioned assumption is experimentally validated in Section V-A. Considering the heavy-tailed characteristics observed in the histograms of the noise samples (see Figure 2), the LAD regression [50] is adopted. By doing so and leveraging the robustness of the ℓ_1 norm against outliers or assumption violations, it is demonstrated that the proposed framework offers a more suitable solution to address the challenges associated with ENF estimation in audio recordings. The estimation of ENF solves the optimization problem

$$\left\{\hat{f}_c, \mathrm{LAD}, \hat{\boldsymbol{\theta}}_{\mathrm{LAD}}\right\} = \underset{f_c,\boldsymbol{\theta}}{\operatorname{argmin}} J_{\mathrm{LAD}}\left(f_c,\, \boldsymbol{\theta}\right), \tag{11}$$

where the objective function of LAD regression is defined as

$$J_{\mathrm{LAD}}(f_c, \boldsymbol{\theta}) = \|\mathbf{x}_{\mathrm{E},l} - \mathbf{H}(f_c)\,\boldsymbol{\theta}\|_1. \tag{12}$$

The objective function defined in (12) is a straightforward extension of the Least Squares (LS) objective function for sinusoidal detection [51], [52]. Here, instead of the LS objective function, the LAD regression is employed. One of the primary strengths of LAD regression is its resilience to outliers (see Section 2.4 in [50]). In (12), $\mathbf{H}(f_c)$ is the linear model used for sinusoidal detection [51], [52]. Its columns $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have elements the cosine at the harmonics of f_c and the sinusoid at the harmonics of f_c. Vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are orthogonal, since $\mathbf{H}(f_c)^\top \mathbf{H}(f_c) = \frac{N_\mathrm{F}}{2}\mathbf{I}$. Specifically, the columns of $\mathbf{H}(f_c)$ are given by:

$$\begin{aligned}\boldsymbol{\alpha} &= \left(1, \cos\left(2\pi\, T, f_c \cdot 1\right), \ldots, \cos\left(2\pi\, T, f_c \cdot (N_\mathrm{F}-1)\right)\right)^\top \\ \boldsymbol{\beta} &= \left(0, \sin\left(2\pi\, T, f_c \cdot 1\right), \ldots, \sin\left(2\pi\, T, f_c \cdot (N_\mathrm{F}-1)\right)\right)^\top,\end{aligned} \tag{13}$$

where $(\cdot)^\top$ stands for the transposition operator. The vector $\boldsymbol{\theta} \in \mathbb{R}^{2\times 1}$ denotes the LAD regression coefficients, serving as weights that dictate how $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ approximate $\mathbf{x}_{\mathrm{E},l}$ so that J_LAD is minimized. $\mathbf{x}_{\mathrm{E},l}$ is a vector with elements $x_{\mathrm{E},l}[n]$, $n \in \{0, 1, \ldots, N_\mathrm{F}-1\}$ denoting the enhanced l-th frame. The optimization problem (11) is solved by estimating f_c and the regression coefficients iteratively. The iteration starts with f_c set at $\hat{f}_{\{\sim\}}[l]$. Then, $\hat{\boldsymbol{\theta}}$ for each frame is found by solving the LAD regression problem [50]:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \, \| \mathbf{x}_{\mathrm{E},l} - \mathbf{H}(\hat{f}_c)\, \boldsymbol{\theta} \, \|_1. \tag{14}$$

The LAD optimization problem with respect to $\boldsymbol{\theta}$ (14) is convex due to the convexity of the ℓ_1 norm. This, combined with linear operations preserving convexity, makes the entire objective function convex. Due to the convexity, the estimates of $\boldsymbol{\theta}$ are computed using the LAD-Least Absolute Shrinkage and Selection Operator (LAD-LASSO), employing the iteratively re-weighted least squares method [50]. Next, $\boldsymbol{\theta}$ is kept fixed and $\hat{f}_c$ for each frame of $\mathbf{x}_{\mathrm{E},l}$, $l \in \{0, 1, \ldots, N_\mathrm{ENF}-1\}$ is estimated. $\hat{f}_{\mathrm{LAD-\{P-MLE\}}}[l]$ minimizes the sum of ℓ_1 norms of approximation errors, i.e.,

$$\hat{f}_{\mathrm{LAD-\{P-MLE\}}}[l] = \underset{f}{\operatorname{argmin}} \sum_{m\in\Omega} \| \mathbf{x}_{\mathrm{E},l} - \mathbf{H}(f)\, \hat{\boldsymbol{\theta}} \|_1, \tag{15}$$

For $\hat{f}_{\mathrm{LAD-\{P-WMLE\}}}[l]$, we are seeking the minimum of the sum of weighted ℓ_1 norms of the approximation error, i.e.,

$$\hat{f}_{\mathrm{LAD-\{P-WMLE\}}}[l] = \underset{f}{\operatorname{argmin}} \sum_{m\in\Omega} w_{m,l} \, \| \mathbf{x}_{\mathrm{E},l} - \mathbf{H}(f)\, \hat{\boldsymbol{\theta}} \|_1, \tag{16}$$

(15) can be adapted to the context of the remaining 8 schemes. The minimization of (15) is conducted by fine grid searching on a dense set comprising a large number of frequency samples. The main contribution is the employment of the ℓ_1 norm for parameter estimation in an alternating optimization (see Algorithm 1) between (14) and (15). The alternating optimization is repeated until the convergence of frequency estimation. Accordingly, although the problem (14) is convex, non-optimality may arise from grid searching in (15).

Algorithm 1 Non-Linear LAD Alternating Optimization

Require: Initial $\hat{f}_c$ and $\boldsymbol{\theta}$, enhanced signal $\mathbf{x}_{\mathrm{E},l}$, maximum iterations T, convergence threshold ϵ, set of harmonics Ω

Ensure: Optimized values $\hat{\boldsymbol{\theta}}$ and $\hat{f}$

```
Initialize iteration count: t = 0
θ̂^(0) = argmin_θ ‖x_{E,l} − H(f̂_c) θ‖_1
while t < T do
    for l = 0 to N_ENF − 1 do
        f̂^(t+1)[l] = argmin_f Σ_{m∈Ω} ‖x_{E,l} − H(f) θ̂^(t)‖_1
    end for
    if |f̂^(t+1) − f̂^(t)| < ε then
        break
    end if
    t = t + 1
end while
return θ̂ and f̂
```

V. EXPERIMENTAL EVALUATION

In this Section, the experimental evaluation of the proposed framework is presented against the state-of-the-art methods. Firstly, it is demonstrated that the Laplacian distribution fits better the histogram of $v_l[n] = x_{\mathrm{E},l}[n] - s_{\mathrm{REF},l}[n], l \in \{0, 1, \ldots, N_\mathrm{ENF}-1\}, n \in \{0, 1, \ldots, N_\mathrm{F}-1\}$ with $s_{\mathrm{REF},l}[n]$ denoting the piecewise constant reference ENF. This observation holds true when compared to the Gaussian distribution, as discussed in Section V-A. Secondly, the effectiveness of the proposed framework is evaluated in Section V-B.

A. STATISTICAL ANALYSIS OF THE DATA

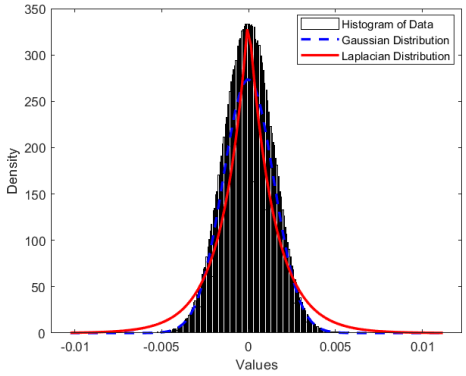
1) DATASET DESCRIPTION

The performance of the proposed framework is evaluated on the benchmark ENF-WHU dataset [20].[2] The ENF-WHU dataset is suitable for ENF estimation tasks due to the inclusion of reference ENF ground truth for evaluating the accuracy of ENF estimation. The dataset consists of 130 real-world audio recordings captured, both day and night, with different weather conditions, around Wuhan University. The duration of audio recordings varies from 5 to 16 minutes. The recordings are mono channel and are resampled at a sampling rate of 8000 Hz with a quantization of 16 bits. Also, a reference dataset of 130 audio recordings is available, containing reference ENF frames sampled at a rate of 400 Hz. Among the audio recordings, 130 audio recordings contain ENF and are organized under the folder `H1`. The remaining 130 recordings contain the reference ENF data and are placed in folder `H1_ref`.

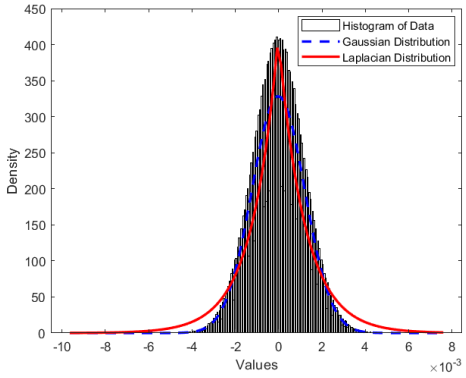
2) EXPERIMENTAL VALIDATION OF MODELING ASSUMPTIONS

In statistical applications, a crucial consideration is determining whether the assumed distribution appropriately represents

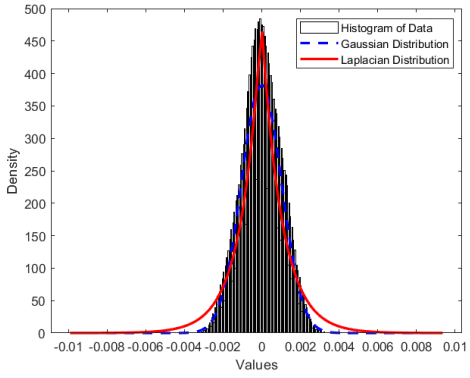
[2]https://github.com/ghua-ac/ENF-WHU-Dataset/tree/master/ENF-WHU-Dataset



(a) Histogram of the noise samples in audio recording #23.



(b) Histogram of the noise samples in audio recording #42.



(c) Histogram of the noise samples in audio recording #129.

FIGURE 2. Histograms of the noise samples $v_l[n]$, $l \in \{0, 1, \ldots, N_{\text{ENF}} - 1\}$ and $n \in \{0, 1, \ldots, N_{\text{F}} - 1\}$ (see Section V) for three randomly chosen audio recordings from the folder `H1` compared to the Laplacian (solid line) and Gaussian (dashed line) probability density functions.

the noise model in the data. This involves examining whether the selected distribution accurately describes the noise model under consideration. The prevailing assumption in state-of-the-art methods postulates a Gaussian distribution for the noise $v_l[n]$ in audio feeds. The proposed framework advocates that the Laplacian distribution fits better than the Gaussian noise distribution.

To corroborate the aforementioned assumption, the histograms of centered noise samples from three randomly chosen audio recordings are depicted in Figure 2. The empirical data distributions seem to fit better the Laplacian distribution (i.e., solid red line) than the Gaussian distribution (i.e., blue dashed line). Table 2 summarizes the results of the Goodness-of-Fit test [53]. Let

TABLE 2. Accepted and rejected hypotheses in Goodness-of-Fit tests for the Laplacian distribution across the 130 noise models pertaining to the 130 audio recordings.

Case	Results
Number of Accepted Hypotheses	87
Number of Rejected Hypotheses	43

$$\breve{u}_l[n] = \frac{u_l[n] - \hat{\mu}_{\tilde{N}}}{\hat{b}_{\tilde{N}}}, \quad l \in \{0, 1, \ldots, N_{\text{ENF}} - 1\} \\ n \in \{0, 1, \ldots, N_{\text{F}} - 1\}, \tag{17}$$

where $\tilde{N} = N_{\text{ENF}} \times N_{\text{F}}$ and $\hat{\mu}_{\tilde{N}}$ is the sample median

$$\hat{\mu}_{\tilde{N}} = \text{med}\left\{ \bigcup_{\ell=0}^{N_{\text{ENF}}-1} \{u_l[0], u_l[1], \ldots u_l[N_{\text{F}} - 1]\} \right\} \tag{18}$$

and $\hat{b}_{\tilde{N}}$ is the MLE of the scale parameter of the Laplacian distribution, i.e., the mean absolute deviation from the median

$$\hat{b}_{\tilde{N}} = \frac{1}{\tilde{N}} \sum_{l=0}^{N_{\text{ENF}}-1} \sum_{n=0}^{N_{\text{F}}-1} \left| u_l[n] - \hat{\mu} \right|. \tag{19}$$

Here, $\mu_{\tilde{N}} \approx 0$, because the data have been centered. The test statistic is defined as

$$Z_{\tilde{N}} = \sqrt{\frac{\tilde{N}}{504}}\, \xi_{\tilde{N}}, \tag{20}$$

where

$$\xi_{\tilde{N}} = \frac{1}{\tilde{N}} \sum_{l=0}^{N_{\text{ENF}}-1} \sum_{n=0}^{N_{\text{F}}-1} (\breve{u}_l[n])^3 - \frac{3}{\tilde{N}^2} \sum_{l=0}^{N_{\text{ENF}}-1} \sum_{n=0}^{N_{\text{F}}-1} \breve{u}_l[n] \\ \cdot \sum_{l=0}^{N_{\text{ENF}}-1} \sum_{n=0}^{N_{\text{F}}-1} (\breve{u}_l[n])^2 + 2\left(\frac{1}{\tilde{N}} \sum_{l=0}^{N_{\text{ENF}}-1} \sum_{n=0}^{N_{\text{F}}-1} \breve{u}_l[n] \right)^2. \tag{21}$$

The asymptotic p-value for the test statistic $Z_{\tilde{N}}$ is calculated as:

$$p\text{-value} = 2 \cdot \left(1 - \Phi\left(\left|Z_{\tilde{N}}\right|\right)\right), \tag{22}$$

where $\Phi(\cdot)$ represents the cumulative distribution function of the standard normal distribution.

Of the 130 hypotheses tested, 87 are accepted, validating the assumption of a Laplacian distribution. However, 43 hypotheses are rejected, suggesting deviation from the Laplacian distribution. To determine the acceptance or rejection of the hypotheses, a significance level of 5% is utilized. Based on the evidence provided in Figure 2 and the results of the Goodness-of-Fit test, it can be concluded that the proposed framework, which utilizes the LAD regression as a consequence and a Laplacian distributed noise model,

TABLE 3. Summary of implementation details of the proposed ENF estimation schemes.

Parameter	Value Description
Sampling Frequency (f_S)	800 Hz
Set of Harmonics ($\mathcal{M}$)	$\{2, 3, 4, 5, 6, 7\}$
Number of Iterations (I)	2
RFA/HRFA Parameter (τ)	2
RFA/HRFA Parameter (α)	$f_S/(4\max_n(\lvert x[n]\rvert))$
Frame Step-size (Δ)	1s (samples)
Frame Length (N_F)	16 f_S (samples)
Search Regions for ENF	$m \in \Omega \subseteq \mathcal{M}$
FFT Frequency Resolution	$1/4000$ Hz
Frequency Scaling	Frequency estimates obtained by STFT are scaled to the 2nd harmonic frequency (100 Hz)

is a suitable choice in the vast majority of audio recordings in [20]. Specifically, for audio recordings #23, #42, and #129, the corresponding p-values are calculated as 0.9016, 0.9812, and 0.7096, respectively. These p-values are all greater than the significance level of 5%, indicating no significant deviation from the assumed Laplacian distribution within these specific noise samples.

B. EXPERIMENTAL SETTINGS AND RESULTS

1) IMPLEMENTATION DETAILS

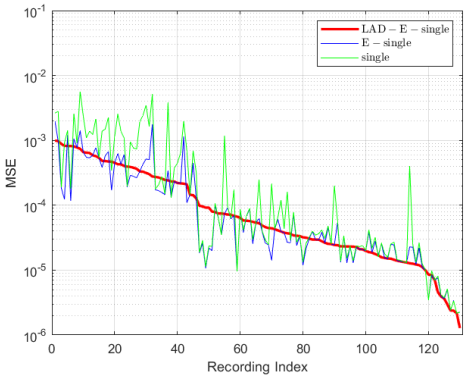
The implementation of the proposed framework is based on the publicly available code[3] for ENF estimation. Following the setup in [20], the implementation details of the proposed END estimation schemes, such as sampling frequency, harmonic set, RFA/HRFA settings, and FFT frequency resolution, are summarized in Table 3. In the context of the RFA/HRFA parameter (α), $x[n]$ denotes the signal composed of a collection of distinct harmonic components derived from the set $\mathcal{M}$. The term $\max_n(\cdot)$ refers to the highest absolute value in the amplitudes of the signal $x[n]$, accounting for both the highest positive and lowest negative amplitudes. The FFT frequency resolution, set to 1/4000, is beneficial when searching for spectral peaks in a dense frequency grid.

2) EVALUATION METRICS

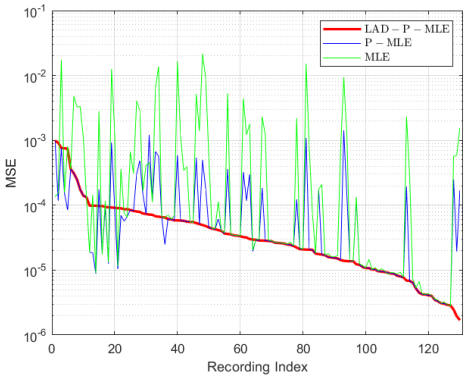
The performance evaluation of the ten proposed ENF estimation schemes against the state-of-the-art (see Table 4) is twofold. First, the results are evaluated by calculating the AMSE, which is the average MSE between the estimated ENF values under `H1` and the ground truth ENF values under `H1_ref` across the 130 recordings. The MSE for each ENF estimation scheme is given by

$$\mathrm{MSE}^{(i)}_{\{\sim\}} = \frac{1}{N_{\mathrm{ENF}}} \sum_{l=0}^{N_{\mathrm{ENF}}-1} \left(f^{(i)}_{\{\sim\}}[l] - f^{(i)}_{\mathrm{REF}}[l] \right)^2, \qquad (23)$$

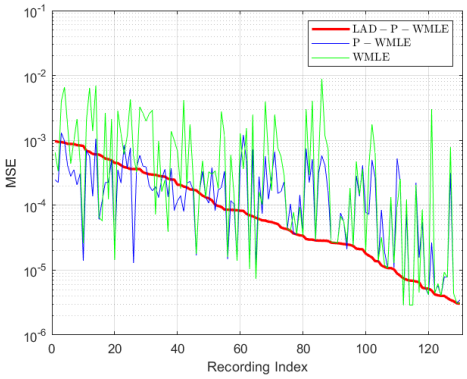
[3]https://github.com/ghua-ac/ENF-WHU-Dataset/tree/master/ENF_Enhancement_Estimation



(a) Comparison of single-tone ENF estimation schemes.



(b) Comparison of MLE ENF estimation schemes.



(c) Comparison of WMLE ENF estimation schemes.

FIGURE 3. Comparison of MSE across the 130 audio recordings from ENF-WHU dataset between the LAD-E-single, LAD-P-MLE, and LAD-P-WMLE ENF estimation schemes and the single/multi-tone competitors.

where $f^{(i)}_{\{\sim\}}$ represents the estimated ENF values of the i-th recording, $i \in \{1, 2, \ldots, 130\}$ and $f^{(i)}_{\mathrm{REF}}$ represents the ground truth ENF values for all frames indexed by length $l \in \{0, 1, \ldots, N_F - 1\}$. In (23), $\{\sim\}$ stands for the different ENF estimation schemes included in Group II, such as the LAD-P-MLE and LAD-P-WMLE. Then, the AMSE across the

TABLE 4. Quantitative performance comparison of the ENF estimation schemes using 130 real-world audio recordings from the ENF-WHU dataset. For single-tone estimation, $\mathcal{M} = \{2\}$. For multi-tone estimation, $\mathcal{M} = \{2, 3, 4, 5, 6, 7\}$.

Group	ENF Estimation Scheme	$M = \lvert\mathcal{M}\rvert$	Harmonic Enhancement	Harmonic Selection	AMSE	std{MSE}
I	single	1	No	No	54×10^{-5}	10×10^{-4}
	E-single [42]	1	RFA	No	20×10^{-5}	3.4×10^{-4}
	AMTC [31]	1	No	No	17×10^{-5}	3.6×10^{-4}
	MLE [21]	6	No	No	140×10^{-5}	37×10^{-4}
	WMLE [22]	6	No	No	95×10^{-5}	16×10^{-4}
	E-MLE [20]	6	HRFA	No	15×10^{-5}	2.9×10^{-4}
	E-WMLE [20]	6	HRFA	No	34×10^{-5}	3.2×10^{-4}
	S-MLE [20]	6	No	GHSA	34×10^{-5}	8.1×10^{-4}
	S-WMLE [20]	6	No	GHSA	28×10^{-5}	7.0×10^{-4}
	P-MLE [20]	6	HRFA	GHSA	13×10^{-5}	2.4×10^{-4}
	P-WMLE [20]	6	HRFA	GHSA	23×10^{-5}	2.5×10^{-4}
II	LAD-single	1	No	No	28×10^{-5}	7.2×10^{-4}
	LAD-E-single	1	RFA	No	18×10^{-5}	2.5×10^{-4}
	LAD-MLE	6	No	No	71×10^{-5}	23×10^{-4}
	LAD-WMLE	6	No	No	68×10^{-5}	14×10^{-4}
	LAD-E-MLE	6	HRFA	No	14×10^{-5}	2.3×10^{-4}
	LAD-E-WMLE	6	HRFA	No	29×10^{-5}	2.6×10^{-4}
	LAD-S-MLE	6	No	GHSA	17×10^{-5}	5.0×10^{-4}
	LAD-S-WMLE	6	No	GHSA	16×10^{-5}	4.3×10^{-4}
	LAD-P-MLE	6	HRFA	GHSA	$\mathbf{9 \times 10^{-5}}$	$\mathbf{1.8 \times 10^{-4}}$
	LAD-P-WMLE	6	HRFA	GHSA	20×10^{-5}	2.0×10^{-4}

130 audio recordings is calculated as:

$$\mathrm{AMSE}_{\{\sim\}} = \frac{1}{130} \sum_{i=1}^{130} \mathrm{MSE}_{\{\sim\}}^{(i)}. \tag{24}$$

The standard deviation std{·} of the calculated MSE is also measured, i.e.,

$$\mathrm{std}\{\cdot\} = \sqrt{\frac{1}{130} \sum_{i=1}^{130} \left(\mathrm{MSE}_{\{\sim\}}^{(i)} - \mathrm{AMSE}_{\{\sim\}}\right)^2}. \tag{25}$$

AMSE proves to be a suitable metric for audio authentication as it captures the sum of squared frequency estimation errors and provides a measure of the alignment between the estimated ENF and the reference ENF.

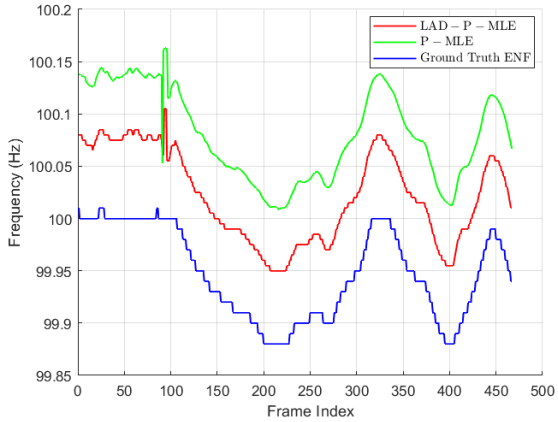
3) EXPERIMENTAL FINDINGS

The performance of the proposed ENF framework across the 130 real-world ENF-WHU audio recordings is evaluated. The implemented methods are presented in Table 4 and categorized into two groups. Group I lists the state-of-the-art single/multi-tone ENF estimation schemes employing harmonic enhancement and harmonic selection modules. Group II presents the evaluation results of the LAD-based ENF estimation framework where RFA/HRFA or GHSA are integrated for single/multi-tone harmonic enhancement and harmonic selection.

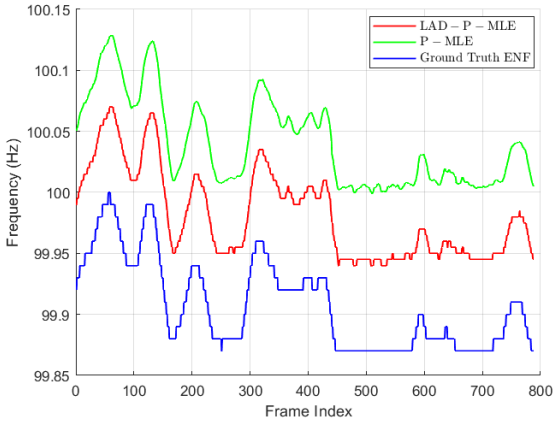
The metrics for ENF estimation schemes in Group II outperform all the corresponding ENF estimation schemes in Group I. The top metrics are indicated in boldface. Among the single-tone ENF estimation methods, where the second harmonic is employed, the LAD-E-single scheme achieves a significantly lower AMSE and std{·} against the two state-of-the-art single and E-single competitors in Group I. More specifically, the evaluation metrics of the proposed LAD-E-single are 18×10^{-5} and 2.5×10^{-4}, respectively. When multiple harmonics are included, the LAD-P-MLE and the LAD-P-WMLE result in a higher ENF estimation accuracy. In Group II, the proposed LAD-P-MLE method outperforms the P-MLE ENF estimation scheme, achieving a lower AMSE and std{·} of 9×10^{-5} and 1.8×10^{-4} compared to 13×10^{-5} and 2.4×10^{-4} for P-MLE, respectively. In Group II, the LAD-P-WMLE scheme has a higher AMSE and std{·} compared to the LAD-P-MLE, suggesting that the LAD-P-MLE performs better within this group. However, it is important to note that even though the LAD-P-WMLE is worse than the LAD-P-MLE in Group II, it still outperforms its direct competitor, P-WMLE, in Group I, as it achieves a lower AMSE and std{·} of 20×10^{-5} and 2.0×10^{-4}, respectively.

By comparing the resulting AMSE values between the extracted ENF from the audio recordings and the reference ENF in Table 4, the legitimacy of the audio signals is determined. A significantly lower AMSE suggests a closer match between the extracted ENF and the reference one, indicating a higher level of authenticity.

In Figure 3, the MSE measured for the three most representative subsets of ENF estimation schemes using all the recordings from the ENF-WHU dataset is presented. The red curve in Figure 3(a), (b), and (c) depicts the MSE values of the proposed LAD-E-single, LAD-P-MLE, and LAD-P-WMLE ENF estimators. One might expect that the



(a) Comparison of the estimated ENF against the ground truth for each frame of the audio recording #1.



(b) Comparison of the estimated ENF against the ground truth for each frame of the audio recording #41.

FIGURE 4. Comparison between the estimated ENF signals using the P-MLE and LAD-P-MLE methods against the ground truth ENF for two randomly selected audio recordings within the `H1` folder. The plots have been shifted vertically by ± 0.05 Hz around the 2nd harmonic.

single-tone schemes would estimate the ENF with higher accuracy compared to the multi-tone ones, given that they only extract a frequency around the 2nd harmonic. However, as can be seen from the red curves in Figure 3(b) and (c), the effect of GHSA and HRFA is remarkable, providing a strong advantage against the single-tone schemes. Among the multi-tone schemes from Figure 3(b), LAD-P-MLE is the most accurate framework. As a result, after the graph-based harmonic selection, the harmonic enhancement of the components in higher harmonics of the signal, and the LAD regression, the MSE is substantially reduced.

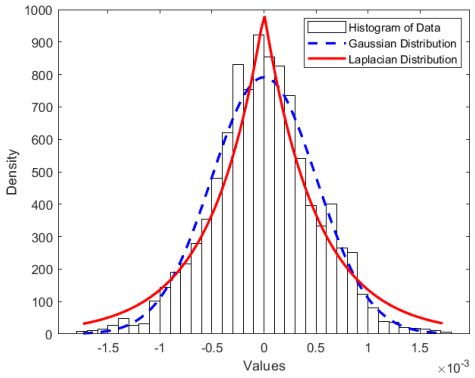
The estimated ENF using either the LAD-P-MLE or P-MLE methods is compared to the ground truth ENF for all frames on two randomly selected audio recordings in Figure 4. To aid visualization, the ENF signals have been shifted along the vertical axis by ± 0.05 Hz. The frame index is indicated in the horizontal axis. It is apparent that the red curve, corresponding to the ENF estimated by the LAD-P-MLE method, exhibits a stronger correlation with the ground truth ENF compared to the ENF estimated by the P-MLE method. This suggests that the LAD-P-MLE method provides more accurate estimates of the ENF for the given audio recordings, making it a more reliable approach in this context. Quantitatively, the calculated MSE between the ENF estimates of the LAD-P-MLE scheme and the ground truth ENF resulted in 6.077×10^{-5} for the audio recording #1, whereas the MSE for the P-MLE method is 8.45×10^{-5}. Similarly, for the audio recording #41, the MSE for the LAD-P-MLE method is calculated as 6.77×10^{-5}, while for the P-MLE method, it is 7.37×10^{-5}. The lower MSE values obtained by the LAD-P-MLE method further confirm its superior accuracy over the P-MLE method in estimating the ENF signal for the provided audio recordings.

4) STATISTICAL ANALYSIS OF THE APPROXIMATION ERROR

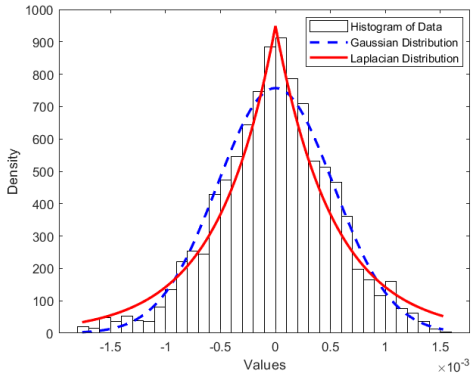
In addition to the evidence provided in Section V-A2, a comprehensive Goodness-of-Fit test [53] is performed to assess the approximation error between $\mathbf{x}_{\mathrm{E},l}$ and $\mathbf{H}(f_c)\boldsymbol{\theta}$ in (14) for the LAD-P-MLE ENF estimation scheme, which is the most accurate framework in Table 4. The test is conducted for the first 5 audio recordings from the folder `H1`, resulting in a total of 2873 frames. The size of the samples to be used is $\tilde{N} = N_{\mathrm{F}}$, where $N_{\mathrm{F}} = 16 \times f_S = 16 \times 800 = 12800$. The objective of the test is to determine whether the null hypothesis that the observed residuals follow the Laplacian distribution is valid.

Within LAD-P-MLE, the LAD-LASSO algorithm is employed, a variant of the LASSO algorithm, where LAD is used as the loss function. LAD-LASSO solves the optimization problem (14). The LAD-LASSO algorithm [50] is justified by the assumption that the approximation error follows the Laplacian distribution. If the centered approximation errors follow the Laplacian distribution, then the ℓ_1 norm penalty in the optimization problem can be interpreted as a Maximum a Posteriori (MAP) prior. The Laplacian distribution is a symmetric density function with a sharp peak at zero and heavy tails, which makes it a suitable choice for modeling data against outliers. The ℓ_1 norm penalty in the LAD-LASSO algorithm is also known as the "Laplace prior" in Bayesian statistics, and it encourages sparse solutions by inducing a "spike-and-slab" prior on the coefficients, where some coefficients are set exactly to zero [54].

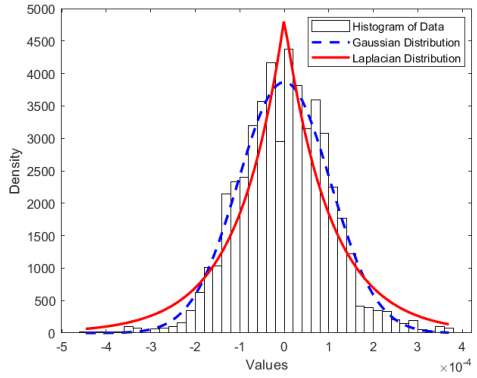
The results of a Goodness-of-Fit test are displayed in Table 5, indicating the number of hypotheses that are accepted versus the number that are rejected. The Laplacian distribution is accepted at a significance level of 5%. From the total 2873 frames tested, the null hypothesis of a Laplacian distribution is accepted for 2620 frames. However, the null hypothesis is rejected in 253 frames. In Figure 5, the histograms of the approximation error for randomly chosen frames are plotted and compared to the Laplacian and Gaussian probability density functions. The error is calculated for three randomly selected frames among the



(a) Histogram of the approximation error for the frame #649.



(b) Histogram of the approximation error for the frame #711.



(c) Histogram of the approximation error for the frame #1633.

FIGURE 5. Histograms of the approximation error between $\mathbf{x}_{E,l}$ and $\mathbf{H}(f_c)\,\theta$ for 3 randomly chosen frames in the first 5 audio recordings under the folder `H1` compared to the Laplacian density function (solid line) and Gaussian density function (dashed line). The approximation errors have been centered.

2873 frames employing the proposed LAD-P-MLE ENF estimation method. The histograms of the approximation error fit better the Laplacian distribution depicted by the red curve (solid line) than the Gaussian distribution. The p-values calculated for frames #649, #711, and #1633 are 0.7827, 0.9563, and 0.9215, respectively. Because the computed p-values are greater than the significance level of 5%, there is no sufficient evidence to warrant the rejection of the claim that the approximation error follows the Laplacian distribution.

TABLE 5. Comparison of accepted and rejected hypotheses for the approximation error in all frames of the first five audio recordings using a Goodness-of-Fit test for the Laplacian distribution.

Case	Results
Number of Accepted Hypotheses	2620
Number of Rejected Hypotheses	253

VI. CONCLUSION AND FUTURE WORK

Ten novel ENF estimation schemes have been proposed. They alternate between LAD regression for finding the regression weights and minimization of objective functions with respect to frequency adapting those in [20], [21], and [22] in the context of the ℓ_1 norm or sum of the ℓ_1 norms of the approximation error. These ENF estimation schemes have enhanced the ability to accurately estimate the ENF against the state-of-the-art, enabling thus the authentication of the legitimacy of audio evidence. The novel ENF estimation schemes have been thoroughly assessed with respect to the average MSE and the standard deviation of MSE using benchmark real audio recordings and reference data. The proposed ENF estimation schemes outperform the state-of-the-art schemes in [20]. Experimental evidence related to Goodness-of-Fit tests has been disclosed to support the validity of the assumption that the noise follows the Laplacian probability density function, justifying the use of LAD regression and ℓ_1 norm-based frequency optimization. These findings lay the foundation for future research dedicated to exploring and advancing further robust spectral analysis methods.

REFERENCES

[1] A. J. Cooper, "An automated approach to the electric network frequency (ENF) criterion—Theory and practice," *Int. J. Speech, Lang. Law*, vol. 16, no. 2, pp. 193–218, Apr. 2010.
[2] D. Nagothu, Y. Chen, A. Aved, and E. Blasch, "Authenticating video feeds using electric network frequency estimation at the edge," *EAI Endorsed Trans. Secur. Saf.*, vol. 7, no. 24, p. e4, Feb. 2021.
[3] M. Savari, A. W. A. Wahab, and N. B. Anuar, "High-performance combination method of electric network frequency and phase for audio forgery detection in battery-powered devices," *Forensic Sci. Int.*, vol. 266, pp. 427–439, Sep. 2016.
[4] C. Grigoras, "Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis," *Forensic Sci. Int.*, vol. 167, nos. 2–3, pp. 136–145, Apr. 2007.
[5] N. Fechner and M. Kirchner, "The humming hum: Background noise as a carrier of ENF artifacts in mobile device audio recordings," in *Proc. 8th Int. Conf. IT Secur. Incident Manag. IT Forensics*, May 2014, pp. 3–13.
[6] S. Vatansever and A. E. Dirik, "Forensic analysis of digital audio recordings based on acoustic mains hum," in *Proc. 24th Signal Process. Commun. Appl. Conf. (SIU)*, May 2016, pp. 1285–1288.
[7] H. Wu, H.-C. Kuo, N. Zheng, K.-H. Hung, H.-Y. Lee, Y. Tsao, H.-M. Wang, and H. Meng, "Partially fake audio detection by self-attention-based fake span discovery," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 9236–9240.
[8] Z. Lv, S. Zhang, K. Tang, and P. Hu, "Fake audio detection based on unsupervised pretraining models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 9231–9235.
[9] D. Nagothu, R. Xu, Y. Chen, E. Blasch, and A. Aved, "Deterring deepfake attacks with an electrical network frequency fingerprints approach," *Future Internet*, vol. 14, no. 5, p. 125, Apr. 2022.
[10] N. Poredi, D. Nagothu, Y. Chen, X. Li, A. Aved, E. Ardiles-Cruz, and E. Blasch, "Robustness of electrical network frequency signals as a fingerprint for digital media authentication," in *Proc. IEEE 24th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2022, pp. 1–6.

[11] A. T. Ho and S. Li, *Handbook of Digital Forensics of Multimedia Data and Devices*. Hoboken, NJ, USA: Wiley, Sep. 2015.

[12] 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*, Nov. 2011, pp. 23–32.

[13] L. Zheng, Y. Zhang, C. E. Lee, and V. L. L. Thing, "Time-of-recording estimation for audio recordings," *Digit. Invest.*, vol. 22, pp. 115–126, Aug. 2017.

[14] S. Vatansever, A. E. Dirik, and N. Memon, "ENF based robust media time-stamping," *IEEE Signal Process. Lett.*, vol. 29, pp. 1963–1967, 2022.

[15] R. Garg, A. Hajj-Ahmad, and M. Wu, "Geo-location estimation from electrical network frequency signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 2862–2866.

[16] C. W. Wong, A. Hajj-Ahmad, and M. Wu, "Invisible geo-location signature in a single image," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1987–1991.

[17] E. Ngharamike, L.-M. Ang, K. P. Seng, and M. Wang, "ENF based digital multimedia forensics: Survey, application, challenges and future work," *IEEE Access*, vol. 11, pp. 101241–101272, 2023.

[18] G. Hua, G. Bi, and V. L. L. Thing, "On practical issues of electric network frequency based audio forensics," *IEEE Access*, vol. 5, pp. 20640–20651, 2017.

[19] A. Hajj-Ahmad, C.-W. Wong, S. Gambino, Q. Zhu, M. Yu, and M. Wu, "Factors affecting ENF capture in audio," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 277–288, Feb. 2019.

[20] G. Hua, H. Liao, H. Zhang, D. Ye, and J. Ma, "Robust ENF estimation based on harmonic enhancement and maximum weight clique," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3874–3887, 2021.

[21] D. Bykhovsky and A. Cohen, "Electrical network frequency (ENF) maximum-likelihood estimation via a multitone harmonic model," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 5, pp. 744–753, May 2013.

[22] A. Hajj-Ahmad, R. Garg, and M. Wu, "Spectrum combining for ENF signal estimation," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 885–888, Sep. 2013.

[23] C. Korgialas and C. Kotropoulos, "Electric network frequency detection using least absolute deviations," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[24] B. Gerazov, Z. Kokolanski, G. Arsov, and V. Dimcev, "Tracking of electrical network frequency for the purpose of forensic audio authentication," in *Proc. 13th Int. Conf. Optim. Electr. Electron. Equip. (OPTIM)*, May 2012, pp. 1164–1169.

[25] Z. Lv, Y. Hu, C.-T. Li, and B.-B. Liu, "Audio forensic authentication based on MOCC between ENF and reference signals," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, Jul. 2013, pp. 427–431.

[26] G. Hua, Y. Zhang, J. Goh, and V. L. L. Thing, "Audio authentication by exploring the absolute-error-map of ENF signals," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 5, pp. 1003–1016, May 2016.

[27] Y. Liu, Z. Yuan, P. N. Markham, R. W. Conners, and Y. Liu, "Application of power system frequency for digital audio authentication," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 1820–1828, Oct. 2012.

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

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

[30] M. El Helou, A. W. Turkmani, R. Chanouha, and S. Charbaji, "A novel ENF extraction approach for region-of-recording identification of media recordings," *Forensic Sci. Int.*, vol. 155, nos. 2–3, p. 165, 2005.

[31] Q. Zhu, M. Chen, C.-W. Wong, and M. Wu, "Adaptive multi-trace carving for robust frequency tracking in forensic applications," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1174–1189, 2021.

[32] L. Fu, P. N. Markham, R. W. Conners, and Y. Liu, "An improved discrete Fourier transform-based algorithm for electric network frequency extraction," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 7, pp. 1173–1181, Jul. 2013.

[33] G. Karantaidis and C. Kotropoulos, "Efficient Capon-based approach exploiting temporal windowing for electric network frequency estimation," in *Proc. IEEE 29th Int. Workshop Mach. Learn. for Signal Process. (MLSP)*, Oct. 2019, pp. 1–6.

[34] A. Triantafyllopoulos, I. Krilis, A. Foliadis, and A. Skodras, "A Hilbert-based approach to the ENF extraction problem," in *Proc. IEICE Inf. Commun. Technol. Forum*, Oct. 2016, pp. 1–6.

[35] G.-O. Glentis, "A fast algorithm for APES and Capon spectral estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4207–4220, Sep. 2008.

[36] G.-O. 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.

[37] G. Karantaidis and C. Kotropoulos, "Blackman–Tukey spectral estimation and electric network frequency matching from power mains and speech recordings," *IET Signal Process.*, vol. 15, no. 6, pp. 396–409, Aug. 2021.

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

[39] A. Hajj-Ahmad, R. Garg, and M. Wu, "Instantaneous frequency estimation and localization for ENF signals," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, Dec. 2012, pp. 1–10.

[40] H. Su, R. Garg, A. Hajj-Ahmad, and M. Wu, "ENF analysis on recaptured audio recordings," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 3018–3022.

[41] X. Lin and X. Kang, "Robust electric network frequency estimation with rank reduction and linear prediction," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 4, pp. 1–13, Oct. 2018.

[42] G. Hua and H. Zhang, "ENF signal enhancement in audio recordings," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1868–1878, 2020.

[43] G. Hua, H. Liao, Q. Wang, H. Zhang, and D. Ye, "Detection of electric network frequency in audio recordings—From theory to practical detectors," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 236–248, 2021.

[44] H. Liao, G. Hua, and H. Zhang, "ENF detection in audio recordings via multi-harmonic combining," *IEEE Signal Process. Lett.*, vol. 28, pp. 1808–1812, 2021.

[45] S. Vatansever, A. E. Dirik, and N. Memon, "Detecting the presence of ENF signal in digital videos: A superpixel-based approach," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1463–1467, Oct. 2017.

[46] P. M. G. I. Reis, J. P. C. da Costa, R. K. Miranda, and G. Del Galdo, "Audio authentication using the kurtosis of ESPRIT based ENF estimates," in *Proc. 10th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Dec. 2016, pp. 1–6.

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

[48] *P1-Policy 1: Load-Frequency Control and Performance*, ENTSOE, UCTE Operations Handbook, Mar. 2009.

[49] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "A tutorial on clique problems in communications and signal processing," *Proc. IEEE*, vol. 108, no. 4, pp. 583–608, Apr. 2020.

[50] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, Oct. 2018.

[51] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[52] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[53] A. Batsidis, P. Economou, and S. Bar-Lev, "A comparative study of goodness-of-fit tests for the Laplace distribution," *Austrian J. Statist.*, vol. 51, no. 2, pp. 91–123, Jan. 2022.

[54] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, Jan. 1996.

CHRISTOS KORGIALAS (Graduate Student Member, IEEE) received the B.Sc. degree in physics from the University of Patras, Greece, in 2020, and the M.Sc. degree in digital media–computational intelligence from the Informatics Department, Aristotle University of Thessaloniki, Greece, in 2022, where he is currently pursuing the Ph.D. degree with the Department of Informatics, Artificial Intelligence and Information Analysis Laboratory. His research focuses on multimodal information analysis, multimedia forensics, signal processing, and machine learning.

CONSTANTINE KOTROPOULOS (Senior Member, IEEE) received the Diploma degree (Hons.) in electrical engineering and the Ph.D. degree in electrical and computer engineering from the Aristotle University of Thessaloniki, in 1988 and 1993, respectively. He is a Full Professor with the Department of Informatics, Aristotle University of Thessaloniki. He was a Visiting Research Scholar with the Department of Electrical and Computer Engineering, University of Delaware, USA, from 2008 to 2009. He has conducted research with the Signal Processing Laboratory, Tampere University of Technology, Finland, in Summer 1993. He has coauthored 66 journal articles and 215 conference papers and contributed nine chapters to edited books in his areas of expertise. He is a Co-Editor of the book *Nonlinear Model-Based Image/Video Processing and Analysis* (J. Wiley and Sons, 2001). His current research interests include forensics; audio, speech, and language processing; signal processing; pattern recognition; multimedia information retrieval; biometric authentication techniques; and human-centered multimodal computer interaction. He was a Scholar of the State Scholarship Foundation, Greece, and the Bodossaki Foundation. He is a member of EURASIP, IAPR, and the Technical Chamber of Greece. He served as the Track Chair for Signal Processing in the 6th International Symposium on Communications, Control, and Signal Processing, Athens, in 2014; the Program Co-Chair for the 4th International Workshop on Biometrics and Forensics (IWBF 2016), Limassol, Cyprus, in 2016; the Program Committee Chair for the XXV European Signal Processing Conference, Kos, Greece, in 2017; the Technical Program Chair for the 5th IEEE Global Conference Signal and Information Processing, Montreal, Canada, in 2017, and the 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing, Rhodes, Greece; and the General Chair for the 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop, Nafplio, Greece. He was a Senior Area Editor of the IEEE Signal Processing Letters. He has been an Editorial Board Member of *Advances in Multimedia*, *International Scholar Research Notices*, *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, *Artificial Intelligence Review*, *Journal of Imaging* (MDPI), *Signals* (MDPI), and *Methods and Protocols* (MDPI).

KONSTANTINOS N. PLATANIOTIS (Fellow, IEEE) received the B.Eng. degree in computer engineering from the University of Patras, Greece, and the M.S. and Ph.D. degrees in electrical engineering from the Florida Institute of Technology, Melbourne, FL, USA. He has been holding the Bell Canada Endowed Chair in Multimedia, since 2014. He is a Professor with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, where he directs the Multimedia Laboratory. His research interests include image/signal processing, machine learning, adaptive learning systems, visual data analysis, multimedia and knowledge media, and affective computing. He is a fellow of the Engineering Institute of Canada and the Canadian Academy of Engineering. He was the Technical Co-Chair of the IEEE 2013 International Conference on Acoustics, Speech and Signal Processing. He served as the Inaugural IEEE Signal Processing Society Vice President for Membership, from 2014 to 2016. He served as the General Co-Chair for the 2017 IEEE GLOBALSIP; the 2018 IEEE International Conference on Image Processing (ICIP 2018); and the IEEE International Acoustics, Speech and Signal Processing (ICASSP 2021). He will be the General Chair of the 2027 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2027), Toronto. He has served as the Editor-in-Chief for the IEEE Signal Processing Letters. He is a registered Professional Engineer in the province of Ontario.