Detection of Electric Network Frequency in Audio Recordings–From Theory to Practical Detectors

Guang Hua, *Member, IEEE*, Han Liao, Qingyi Wang, Haijian Zhang, *Member, IEEE*, and Dengpan Ye, *Member, IEEE*

***Abstract*— Recently, it has been discovered that the electric network frequency (ENF) could be captured by digital audio, video, or even image files, and could further be exploited in forensic investigations. However, the existence of the ENF in multimedia content is not a sure thing, and if the ENF is not present, ENF-based forensic analysis would become useless or even misleading. In this paper, we address the problem of ENF detection in digital audio recordings, which is modeled as the detection of a weak (ENF) signal contaminated by unknown colored wide-sense stationary (WSS) Gaussian noise, while the signal also contains multiple unknown random parameters. We first derive three Neyman-Pearson (NP) detectors, i.e., general matched filter (GMF), matched filter (MF)-like detector, and the asymptotic approximation of the GMF, and choose the MF-like detector as the clairvoyant detector. For practical detectors, we show that the generalized likelihood ratio test (GLRT) could not be efficiently obtained due to the unknown noise and large matrix inversion. Alternatively, we propose two least-squares (LS)-based time domain detectors termed as LS-likelihood ratio test (LRT) and naive-LRT. Further, we propose a time-frequency (TF) domain detector, termed as TF detector, which exploits the *a priori* knowledge of the ENF. The performances of the derived detectors are extensively analyzed in terms of test statistic distributions, threshold selection, and computational complexity. The naive-LRT detector is found to be only effective for very short recordings. As the data recording length increases, both LS-LRT and TF detectors yield effective detection results, while the latter is approximately a constant false alarm rate (CFAR) detector. Practical experiments using real audio recordings justify the effectiveness of the proposed detectors and our analysis.**

***Index Terms*— Digital forensics, electric network frequency, ENF detection, GLRT detector, signal detection.**

I. Introduction

THE electric network frequency (ENF), which is the power transmission frequency having a nominal value at 50 or 60 Hz, has been used as an important forensic criterion, thanks to its properties of random fluctuation around the nominal value and intra-grid consistency [1]. Recently, it has been discovered that the ENF could be captured by not only audio recordings [1], but also video recordings [2], or even a single image [3], and this has significantly widened the applicability of the ENF criterion in forensic investigations.

During an ENF-based forensic investigation process, the ENF signal is first extracted from the questioned file. Then, with or without a reference ENF segment that may potentially correspond to the questioned file in time, one could perform examinations including time-of-recording (timestamp) verification [4]–[6], tampering detection [7]–[12], camera forensics [13], and region of recording estimation [3], [14], [15], etc.

Despite the extensive prior works on ENF estimation [16]–[21] and the above mentioned applications, a very fundamental and important problem, i.e., ENF detection, has not been effectively addressed in the literature. It is known that emitted from AC driven appliances and devices, the ENF exists as an acoustic hum and time-varying light intensity, which could respectively be captured by nearby microphones and cameras. However, since the nominal frequency (as well as several harmonic frequencies) is in very low frequency band, normal movement of the recording device could cause strong Doppler effects and destroy the ENF signal in audio and video content [2], [22]. Besides, the subtle ENF signal is very weak compared to the interested multimedia content. For audio recording, ENF strength is inversely proportional to the squared distance between the microphone and the nearest ENF source, while for video recording, light intensity inconsistency is usually hardly noticeable. The general flowchart of an ENF-based audio forensic examination system is depicted in Fig. 1 where the ENF detection module is marked gray, and the dashed arrows indicate that the corresponding modules are optional. If a questioned file did not capture the ENF signal or the captured ENF signal is severely corrupted by interference, Doppler effects, or noise, then ENF-based forensic analysis would become useless. Even worse, any further analysis implicitly presuming successful capture of the ENF signal would mislead related investigations. For example, the existing procedures for ENF-based audio timestamp verification (e.g., [1], [4], [6], etc.) could still be carried out for a recording without ENF (we do not perform ENF detection and presume its existence), and we could still obtain a result of "matched" timestamp. However, adopting such a result would

Manuscript received December 26, 2019; revised April 19, 2020; accepted July 6, 2020. Date of publication July 17, 2020; date of current version July 31, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61802284, Grant U1636101, Grant U1736211, and Grant U1636219, in part by the National Key Research Development Program of China under Grant 2019QY(Y)0206 and Grant 2016QY01W0200, and in part by the Hubei Provincial Natural Science Foundation of China under Grant 2018CFB225 and Grant 2019CFB512. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrick Bas. *(Corresponding author: Haijian Zhang.)*

Guang Hua, Han Liao, Qingyi Wang, and Haijian Zhang are with the School of Electronic Information, Wuhan University, Wuhan 430072, China (e-mail: ghua@whu.edu.cn; liaohan@whu.edu.cn; wqywqy@whu.edu.cn; haijian.zhang@whu.edu.cn).

Dengpan Ye is with the School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China (e-mail: yedp@whu.edu.cn).

Digital Object Identifier 10.1109/TIFS.2020.3009579

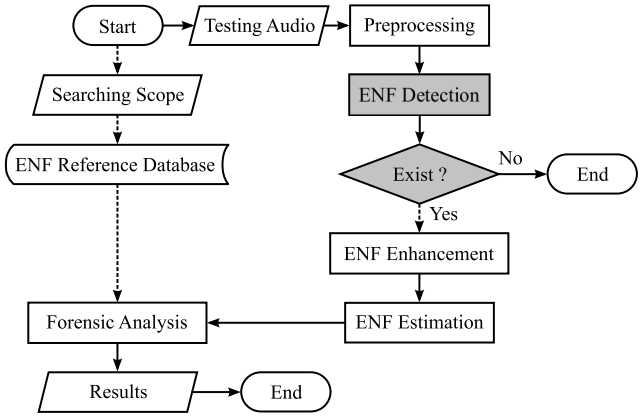


Fig. 1. General flowchart of an ENF-based audio forensic examination system and the relationship to the focus of this paper, i.e., ENF detection.

definitely lead us to a wrong decision, since it makes no sense to base the forensic analysis on a destroyed or nonexistent signal.

Therefore, as shown in Fig. 1, it is essential for us to gain knowledge of whether the ENF exists in a questioned file before conducting further forensic analysis based on it. In [3] and [23], the authors have addressed the ENF detection problem in video and image respectively, while in this paper, we focus on ENF detection in audio recordings.

In Fig. 1, the preprocessing step downsamples and bandpass filters the audio file under investigation, which respectively eliminates redundant computational loads and suppresses out-of-band noise and interference. Therefore, ENF detection in audio recordings is modeled as the detection of the weak ENF signal contaminated by unknown colored (due to bandpass filtering) wide-sense stationary (WSS) Gaussian noise. The problem is generally difficult not only because of the ENF being weak, but also because the amplitudes and instantaneous frequencies (IFs) are random processes, and the initial phase of the ENF signal is a random variable.

To tackle this problem, we first assume that the ENF signal is known and derive three Neyman-Pearson (NP) detectors, i.e., general matched filter (GMF), matched filter (MF)-like detector, and the asymptotic approximation of the GMF, and choose the MF-like detector as the clairvoyant detector for performance bound. For practical detectors, making use of the slowly varying nature of the ENF, we consider the amplitudes, the ENF, as well as the initial phase, all as unknown constants and formulate a generalized likelihood ratio test (GLRT). Although it is found that the GLRT could not be practically obtained due to the unknown noise, we alternatively propose two least-squares (LS)-based time domain detectors termed as LS-LRT and naive-LRT. Further, we propose a time-frequency (TF) domain detector, termed as TF detector, which exploits the *a priori* knowledge of the ENF gained from widely available reference ENF databases. The TF detector performs short-time Fourier transform (STFT)-based TF analysis of the audio signal. Then it uses the sample variance of the IFs as the test statistic. Via extensive quantitative analysis of the test statistics, it is then shown that the decision thresholds could be effectively set without knowing the noise parameter for LS-LRT, naive-LRT, and TF detectors. Further, the TF detector is approximately a constant false alarm rate (CFAR) detector. The effectiveness of the proposed detectors is supported by both synthetic simulations and practical experiments using real audio recordings. It is found that the naive-LRT detector is only effective for very short recordings. As the data recording length increases, both LS-LRT and TF detectors yield effective and competitive detection results.

The paper is organized as follows. General signal model and problem formulation are provided in Section II. In Section III, based on different assumptions and approximations of the detection problem, we present the detailed derivations of the clairvoyant and practical detectors. In Section IV, extensive performance analysis is carried out to decide the thresholds and compare the detection performance and computational complexity of the proposed detectors. Experimental results using real audio recordings are presented in Section V. Conclusion is made in Section VI.

II. Signal Model and General Problem

The time-domain audio samples after downsampling, denoted by $\tilde{x}[n]$, is modeled by the superposition of the ENF waveform $\tilde{s}[n]$, intended audio content $\tilde{c}[n]$, and background noise $\tilde{v}[n]$. Thus the original ENF detection problem could be formulated as a binary hypothesis testing

$$\begin{aligned}\mathcal{H}_0 :&\tilde{x}[n] = \tilde{c}[n] + \tilde{v}[n],\\ \mathcal{H}_1 :&\tilde{x}[n] = \tilde{s}[n] + \tilde{c}[n] + \tilde{v}[n],\end{aligned} \tag{1}$$

which is generally a challenging weak signal detection problem. In this problem, the audio content (interference) $\tilde{c}[n]$ is strong and highly non-stationary, and $\tilde{v}[n]$ is assumed to be independent and identically distributed (i.i.d.) white Gaussian noise (WGN), $\forall n$, $\tilde{v}[n] \sim \mathcal{N}(0, \sigma_{\tilde{v}}^2)$. The second stage of the preprocessing step applies a bandpass filter to remove most of the interference and noise. Suppose an ideal bandpass filter with a very narrow passband centered at the nominal frequency[1] is used and the interference $\tilde{c}[n]$ is approximately removed by the filtering process [10], [17], [21], we could then formulate the ENF detection problem after preprocessing ($\{\tilde{\cdot}\}$ removed) as

$$\begin{aligned}\mathcal{H}_0 :&x[n] = v[n],\\ \mathcal{H}_1 :&x[n] = s[n] + v[n],\end{aligned} \tag{2}$$

where $n \in \{0, 1, \ldots, N-1\}$,

$$s[n] \approx \tilde{s}[n] = A[n]\cos\left(2\pi T \sum_{i=0}^{n} f[i] + \phi\right), \tag{3}$$

$A[n] > 0$ and $f[n]$ are the unknown time-varying amplitudes and the ENF respectively, and ϕ is the unknown initial phase.

[1]The passband could be centered at the nominal frequency or at the harmonic(s) according to the specific situation within a given grid. In our theoretical analysis, the passband is centered at 50 Hz while for experiment using practical recordings, the passband is centered at 100 Hz which contains the strongest ENF harmonic in central China grid.

Here $T = 1/f_S$ is the sampling interval, and f_S is the sampling frequency. Here we define the signal-to-noise ratio (SNR) by

$$\mathrm{SNR} = 10 \log \frac{\sum_n \tilde{s}^2[n]}{\sum_n \tilde{v}^2[n]}, \tag{4}$$

based on the original unfiltered noise. The bandpass filtering process reshapes the spectrum of the WGN, making $v[n]$ a band-limited colored Gaussian noise, $\forall n,\ v[n] \sim \mathcal{N}(0, \sigma_v^2)$. Denote the width of the passband as B Hz, then we have $\sigma_v^2 = 2TB\sigma_{\tilde{v}}^2$ [24].

Remark: The problem (2) is the detection of a random signal $s[n]$ contaminated by colored WSS Gaussian noise $v[n]$, while the random signal $s[n]$ contains stochastic processes $A[n]$ and $f[n]$, plus a random variable ϕ. Although the *a priori* information about these unknown parameters is available, i.e., $A[n]$ following i.i.d. Gaussian distribution with unit mean and small variance, $f[n]$ following a first-order auto-regressive (AR) model [25], and ϕ being uniformly distributed, $\phi \sim \mathcal{U}[0, 2\pi]$, the problem (2) is still very challenging [26]. What is worse, we have to deal with the nuisance parameter σ_v^2 of the unknown noise. Therefore, it is necessary for us to explore the performance limit of ENF detector on the one hand, and on the other hand, look into approximations or alternative reformulations of (2) for mathematically tractable and practically implementable solutions.

III. ENF Detectors

A. Clairvoyant Detector

To quantify ENF detection performance limit, we first derive the clairvoyant detector, i.e., the NP detector assuming perfect knowledge of the signal $s[n]$. Note that such a detector is infeasible in practice due to the unknown signal and noise parameters, thus it only serves as a performance upper bound. In this subsection, we consider problem (2) as the detection of a known deterministic signal $s[n]$ in the noisy observation $x[n]$. However, since the noise $v[n]$ is non-white, a GMF is catered for this problem.

1) Generalized Matched Filter: Let the impulse response of the M-order bandpass finite impulse response (FIR) filter be $h[n]$, $n = 0, 1, \ldots, M$, and given the auto-correlation function of $\tilde{v}[n]$ by $\gamma_{\tilde{v}}[m] = E\{\tilde{v}[n]\tilde{v}[n-m]\} = \sigma_{\tilde{v}}^2 \delta[m]$, where $E\{\cdot\}$ denotes expectation, we then have [27]

$$\begin{aligned} \gamma_v[m] &= E\{v[n]v[n-m]\} \\ &= h[-m] \otimes h[m] \otimes \gamma_{\tilde{v}}[m] \\ &= \sigma_{\tilde{v}}^2 \sum\nolimits_{\tau} h[\tau]h[\tau - m], \end{aligned} \tag{5}$$

where $\otimes$ denotes linear convolution. Note that $v[n]$ is a WSS colored Gaussian noise which could be modeled by an M-order moving average (MA) process. According to (5), the $N \times N$ real covariance matrix of $v[n]$ is given by

$$[\mathbf{C}]_{ij} = \gamma_v[i - j], \tag{6}$$

which is a symmetric Toeplitz matrix. The NP detector, i.e., GMF, decides $\mathcal{H}_1$ if the test statistic satisfies [26, pp. 106]

$$T_{\mathrm{GMF}}(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{s} > \eta_{\mathrm{GMF}}, \tag{7}$$

where η_{GMF} is a threshold value, $\mathbf{x} = [x[0], \ldots, x[N-1]]^T$, $\mathbf{s} = [s[0], \ldots, s[N-1]]^T$, and $[\cdot]^T$ is the transpose operator. The test statistics under the two hypotheses follow Gaussian distributions with [26]

$$\begin{cases} E\{T_{\mathrm{GMF}}(\mathbf{x}); \mathcal{H}_0\} = 0, \\ E\{T_{\mathrm{GMF}}(\mathbf{x}); \mathcal{H}_1\} = \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}, \\ \mathrm{var}\{T_{\mathrm{GMF}}(\mathbf{x}); \mathcal{H}_0\} = \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}, \\ \mathrm{var}\{T_{\mathrm{GMF}}(\mathbf{x}); \mathcal{H}_1\} = \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}. \end{cases} \tag{8}$$

Although the above GMF could provide optimal detection performance for deterministic signal corrupted in correlated noise, it suffers from the unknown $\mathbf{C}$ and needs to perform computationally-intensive matrix inversion. To obtain an effective and computationally efficient NP detector, one may resort to two kinds of approximations of the GMF.

2) Matched Filter-Like Approximation: The first approximation acts as if the noise were white and directly takes the inner product of $\mathbf{s}$ and $\mathbf{x}$ as the test statistic. The resulting MF-like detector decides $\mathcal{H}_1$ if [26, pp. 101]

$$T_{\mathrm{MF}}(\mathbf{x}) = \mathbf{x}^T \mathbf{s} > \eta_{\mathrm{MF}}. \tag{9}$$

Although the term $\mathbf{C}$ is omitted in the test statistic, noise is still non-white and the distributions of $T_{\mathrm{MF}}(\mathbf{x})$ under both hypotheses will still be dependent on $\mathbf{C}$. Therefore, similar to (8), we have

$$\begin{cases} E\{T_{\mathrm{MF}}(\mathbf{x}); \mathcal{H}_0\} = 0, \\ E\{T_{\mathrm{MF}}(\mathbf{x}); \mathcal{H}_1\} = \mathbf{s}^T \mathbf{s} = \|\mathbf{s}\|_2^2, \\ \mathrm{var}\{T_{\mathrm{MF}}(\mathbf{x}); \mathcal{H}_0\} = \mathbf{s}^T \mathbf{C} \mathbf{s}, \\ \mathrm{var}\{T_{\mathrm{MF}}(\mathbf{x}); \mathcal{H}_1\} = \mathbf{s}^T \mathbf{C} \mathbf{s}, \end{cases} \tag{10}$$

where the variance term is different from white noise case ($\mathbf{C} \neq \sigma_v^2 \mathbf{I}$) and is obtained via the following derivation

$$\begin{aligned} &\mathrm{var}\{T_{\mathrm{MF}}(\mathbf{x}); \mathcal{H}_0\} \\ &\quad = E\left\{ \left(\mathbf{s}^T \mathbf{v} - E\left\{ \mathbf{s}^T \mathbf{v} \right\} \right)^2 \right\} \\ &\quad = E\left\{ \mathbf{s}^T \mathbf{v} \mathbf{v}^T \mathbf{s} \right\} = \mathbf{s}^T E\left\{ \mathbf{v} \mathbf{v}^T \right\} \mathbf{s} = \mathbf{s}^T \mathbf{C} \mathbf{s}, \end{aligned} \tag{11}$$

where $\mathbf{v} = [v[0], v[1], \ldots, v[N-1]]^T$. It could be verified that the variances under the two hypotheses are identical. To compare (8) and (10), we normalize $T_{\mathrm{GMF}}(\mathbf{x})$ and $T_{\mathrm{MF}}(\mathbf{x})$ such that

$$\frac{T_{\mathrm{GMF}}(\mathbf{x})}{\sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}} \sim \begin{cases} \mathcal{N}(0, 1), & \text{under } \mathcal{H}_0, \\ \mathcal{N}\left(\sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}, 1 \right), & \text{under } \mathcal{H}_1, \end{cases} \tag{12}$$

and

$$\frac{T_{\mathrm{MF}}(\mathbf{x})}{\sqrt{\mathbf{s}^T \mathbf{C} \mathbf{s}}} \sim \begin{cases} \mathcal{N}(0, 1), & \text{under } \mathcal{H}_0, \\ \mathcal{N}\left(\sqrt{\left(\mathbf{s}^T \mathbf{s}\right)^2 / \mathbf{s}^T \mathbf{C} \mathbf{s}}, 1 \right), & \text{under } \mathcal{H}_1. \end{cases} \tag{13}$$

Therefore, the normalized test statistics respectively follow two Gaussian distributions with unit variance. Since $\mathbf{C}$ and $\mathbf{C}^{-1}$ are both positive definite matrices and $\mathbf{s}$ is a non-zero vector, the ***Kantorovich inequality*** holds that

$$\begin{aligned} &\mathbf{s}^T \mathbf{C} \mathbf{s} \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} > \left(\mathbf{s}^T \mathbf{s} \right)^2 \\ &\Rightarrow \sqrt{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} > \sqrt{\left(\mathbf{s}^T \mathbf{s}\right)^2 / \mathbf{s}^T \mathbf{C} \mathbf{s}}, \end{aligned} \tag{14}$$

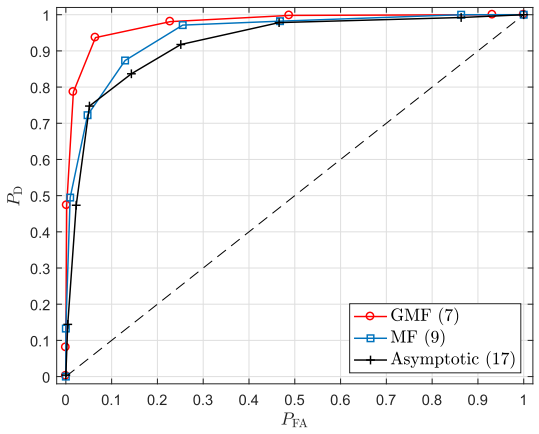


Fig. 2. Performance comparison of the GMF (7), MF (9), and asymptotic approximation (17) using synthetic data, where $f_{\mathrm{S}} = 400$ Hz, signal duration is 5 seconds, and SNR $= -25$ dB.

indicating that the probability density functions (PDFs) under $\mathcal{H}_0$ and $\mathcal{H}_1$ obtained from the GMF are further separated from each other than the ones obtained from the MF approximation. Performance degradation of the MF approximation could hence be quantified. Note that we omitted the equality in (14) because it only holds if $\mathbf{s} = \mathbf{0}$ or $\mathbf{C} = \sigma_v^2\mathbf{I}$, i.e., the bandpass filtered noise is white, which could never happen.

3) Asymptotic Approximation: The second approximation of the GMF relies on the asymptotic property. Define the correlation time of $v[n]$, denoted by T_v, as

$$T_v = \arg\min_{\tau} \tau \quad \text{s.t.} \quad \forall m > \tau, \gamma_v[m] \approx 0. \tag{15}$$

Meanwhile, it follows from (5) that $\forall m > M$, $\gamma_v[m] = 0$, thus we have $T_v = M$. Therefore, if

$$N \gg T_v = M, \tag{16}$$

then the GMF (7) could be approximated by [26, pp. 108]

$$T^{\infty}_{\mathrm{GMF}}(\mathbf{x}) = \int_0^{2\pi} \frac{X(\omega)S^*(\omega)}{P_v(\omega)} d\omega > \eta_{\mathrm{GMF}} \tag{17}$$

to decide $\mathcal{H}_1$, where $\{\cdot\}^*$ denotes complex conjugation, $X(\omega)$ and $S(\omega)$ are the discrete-time Fourier transform (DTFT) of $x[n]$ and $s[n]$ respectively, and $P_v(\omega)$ is the power spectral density (PSD) of $v[n]$. The PDFs of $T^{\infty}_{\mathrm{GMF}}(\mathbf{x})$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ are also Gaussian and they asymptotically follow (8). The asymptotic approximation of the GMF could avoid the calculation of $\mathbf{C}^{-1}$ in the test statistic, providing a trade-off between computational complexity and detection performance.

4) Discussion: Among the three NP detectors, only the MF yields a test statistic independent of the unknown noise parameter. For the GMF and asymptotic approximation, the distributions of the test statistics under $\mathcal{H}_0$ and $\mathcal{H}_1$ are unknown because of $\mathbf{C}^{-1}$. Therefore, the MF detector could effectively determine the centroids of test statistic distributions under $\mathcal{H}_0$ and $\mathcal{H}_1$, and a safe noise independent threshold could be obtained as $\eta_{\mathrm{MF}} = 1/2\|\mathbf{s}\|_2^2$, indicated from (10). While performance degradation of the MF as compared to the GMF is in terms of shifting the test statistic PDF under $\mathcal{H}_1$ slightly closer to the PDF under $\mathcal{H}_0$ (see (12)-(14)), the performance of the asymptotic approximation is affected by the asymptotic condition (16). Fig. 2 shows the receiver operating characteristics (ROC) curves of the three detectors using synthetic data, in which we observe comparable performances between the two approximated detectors and the slight performance degradation against the GMF. Based on the above analysis, the MF detector is the best choice for the clairvoyant detector for its weak dependence on noise parameter, the lowest computation complexity, and competitive detection performance.

B. GLRT Detectors

In this subsection, we consider problem (2) as the detection of an unknown deterministic signal $s[n]$ in the noisy observation $x[n]$ and propose two practical time domain ENF detectors. During the derivations we assume noise parameters are known and look into whether a test statistic and its corresponding distributions are independent of the noise parameters. Exploiting the slowly varying nature of the ENF, we treat $A[n]$ and $f[n]$ as constant in this subsection and rewrite (3) as

$$s_{\mathrm{c}}[n] = A_{\mathrm{c}} \cos\left(2\pi T f_{\mathrm{c}} n + \phi\right), \tag{18}$$

where A_{c} and f_{c} are constant approximations of $A[n]$ and $f[n]$ respectively. Then, (2) becomes

$$\begin{aligned} \mathcal{H}_0 &: x[n] = v[n], \\ \mathcal{H}_1 &: x[n] = s_{\mathrm{c}}[n] + v[n], \end{aligned} \tag{19}$$

and the GLRT detector decides $\mathcal{H}_1$ if

$$\ln \frac{p(\mathbf{x}; \hat{A}_{\mathrm{c}}, \hat{f}_{\mathrm{c}}, \hat{\phi}, \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \eta, \tag{20}$$

where $\hat{A}_{\mathrm{c}}$, $\hat{f}_{\mathrm{c}}$, and $\hat{\phi}$ are the estimates of the unknown parameters under $\mathcal{H}_1$. We first derive the standard GLRT that substitutes the maximum likelihood estimates (MLEs) of the unknown parameters into (20) for solution and then discuss about two more computationally efficient approximations.

Let

$$\begin{cases} \alpha = A_{\mathrm{c}} \cos\phi, \\ \beta = -A_{\mathrm{c}} \sin\phi, \end{cases} \tag{21}$$

and let

$$\begin{cases} \mathbf{a} = [1, \cos\left(2\pi T f_c \cdot 1\right), \ldots, \cos\left(2\pi T f_c (N-1)\right)]^T, \\ \mathbf{b} = [0, \sin\left(2\pi T f_c \cdot 1\right), \ldots, \sin\left(2\pi T f_c (N-1)\right)]^T, \end{cases} \tag{22}$$

then the likelihood function under $\mathcal{H}_1$ is given by

$$\begin{aligned} p(\mathbf{x}; \mathcal{H}_1) = &\frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{C}|^{\frac{1}{2}}} \\ &\times \exp\left[-\frac{[\mathbf{x} - \mathbf{H}(f_{\mathrm{c}})\boldsymbol{\theta}]^T \mathbf{C}^{-1} [\mathbf{x} - \mathbf{H}(f_{\mathrm{c}})\boldsymbol{\theta}]}{2}\right], \end{aligned} \tag{23}$$

where $\boldsymbol{\theta} = [\alpha, \beta]^T$ and $\mathbf{H}(f_{\mathrm{c}}) = [\mathbf{a}, \mathbf{b}]$. In the sequel, $\mathbf{H}(f_{\mathrm{c}})$ is sometimes written as $\mathbf{H}$ for notation clarity and brevity. The MLEs of the unknown parameters are then given by [28]

$$\{\hat{f}_{\mathrm{c}}, \hat{\boldsymbol{\theta}}\} = \arg\min_{f_{\mathrm{c}}, \boldsymbol{\theta}} J(f_{\mathrm{c}}, \boldsymbol{\theta}), \tag{24}$$

where

$$J\left(f_{\mathrm{c}},\boldsymbol{\theta}\right)=\left[\mathbf{x}-\mathbf{H}\left(f_{\mathrm{c}}\right)\boldsymbol{\theta}\right]^{T}\mathbf{C}^{-1}\left[\mathbf{x}-\mathbf{H}\left(f_{\mathrm{c}}\right)\boldsymbol{\theta}\right] \tag{25}$$

is a quadratic function of $\boldsymbol{\theta}$, and it follows that

$$\hat{\boldsymbol{\theta}}=[\mathbf{H}^{T}(f_{\mathrm{c}})\mathbf{C}^{-1}\mathbf{H}(f_{\mathrm{c}})]^{-1}\mathbf{H}^{T}\left(f_{\mathrm{c}}\right)\mathbf{C}^{-1}\mathbf{x}. \tag{26}$$

Substitute (26) into (25), then the MLE of f_{c} corresponds to the minimum value of

$$J(f_{\mathrm{c}},\hat{\boldsymbol{\theta}})=\mathbf{x}^{T}\left[\mathbf{C}^{-1}-\mathbf{C}^{-1}\mathbf{H}(\mathbf{H}^{T}\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^{T}\mathbf{C}^{-1}\right]\mathbf{x}. \tag{27}$$

Since (27) is independent of $\hat{\boldsymbol{\theta}}$, parameter $\hat{f}_{\mathrm{c}}$ could be independently solved and then substituted into (26) to obtain $\hat{\boldsymbol{\theta}}$, and the inverse of (21) yields $\hat{A}_{\mathrm{c}}$ and $\hat{\phi}$.

However, since the nuisance noise parameter $\mathbf{C}$ is unknown, the above MLEs could not be obtained in practice. Therefore, we exploit alternative LS-based detectors.

1) LS-LRT: The LS-LRT obtains the estimates of the unknown variables by minimizing the squared difference between $\mathbf{x}$ and $\mathbf{H}\boldsymbol{\theta}$, i.e.,

$$\{\hat{f}_{\mathrm{c,LS}},\hat{\boldsymbol{\theta}}_{\mathrm{LS}}\}=\arg\min_{f_{\mathrm{c}},\boldsymbol{\theta}}J_{\mathrm{LS}}\left(f_{\mathrm{c}},\boldsymbol{\theta}\right), \tag{28}$$

where

$$J_{\mathrm{LS}}\left(f_{\mathrm{c}},\boldsymbol{\theta}\right)=\left(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\right)^{T}\left(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\right). \tag{29}$$

It then follows that [28, pp. 139]

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}}=(\mathbf{H}^{T}\mathbf{H})^{-1}\mathbf{H}^{T}\mathbf{x}\approx\frac{2}{N}\mathbf{H}^{T}(\hat{f}_{\mathrm{c,LS}})\mathbf{x}, \tag{30}$$

and the LS-based estimation of the ENF corresponds to the peak of the periodogram of $x[n]$,

$$\hat{f}_{\mathrm{c,LS}}=\arg\max_{f}\left|\sum_{n=0}^{N-1}x[n]\exp\left(j2\pi Tfn\right)\right|^{2}. \tag{31}$$

The estimates (30) and (31) are in fact identical to those derived in white noise scenario because (28) becomes equivalent to (24) when $\mathbf{C}=\sigma_{v}^{2}\mathbf{I}$. Fortunately, it has been proven in [29] that the LS-based estimation asymptotically achieves the same statistical performance as the MLE even in the colored noise case. Substitute (30) and (31) into (18) and use the estimated signal as the MF, we thus obtain the practical MF-like detector (similar to (9)) which decides $\mathcal{H}_1$ if

$$\begin{aligned}T_{\mathrm{LS}}(\mathbf{x})&=\frac{\mathbf{x}^{T}\hat{\mathbf{s}}_{\mathrm{LS}}}{\mathbf{x}^{T}\mathbf{x}}=\frac{\mathbf{x}^{T}\mathbf{H}(\hat{f}_{\mathrm{c,LS}})\hat{\boldsymbol{\theta}}_{\mathrm{LS}}}{\mathbf{x}^{T}\mathbf{x}}\\&\approx\frac{2\mathbf{x}^{T}\mathbf{H}(\hat{f}_{\mathrm{c,LS}})\mathbf{H}^{T}(\hat{f}_{\mathrm{c,LS}})\mathbf{x}}{N\mathbf{x}^{T}\mathbf{x}}>\eta_{\mathrm{LS}},\end{aligned} \tag{32}$$

where (30) is used to obtain the approximation.

2) Naive-LRT: The naive-LRT detector directly uses the nominal value of the ENF as the frequency estimate, and we denote it by $\hat{f}_{\mathrm{c,naive}}=50$ Hz (or 100 Hz). In this way, the peak search process in (31) could be avoided. Making use of the *a priori* knowledge of the ENF, the naive-LRT detector thus trades off frequency estimation accuracy for computational efficiency. Replace $\hat{f}_{\mathrm{c,LS}}$ by $\hat{f}_{\mathrm{c,naive}}$, we have

$$\hat{\boldsymbol{\theta}}_{\mathrm{naive}}\approx\frac{2}{N}\mathbf{H}^{T}(\hat{f}_{\mathrm{c,naive}})\mathbf{x}, \tag{33}$$

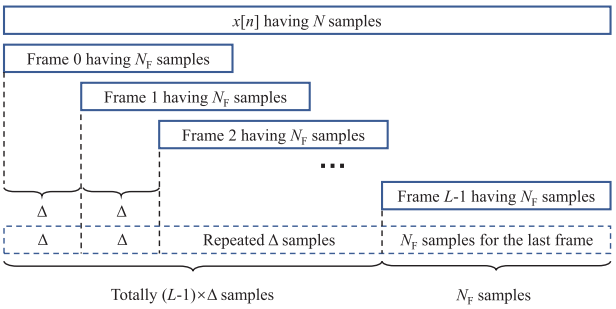


Fig. 3. Demonstration of frame-based processing.

and the naive-LRT detector decides $\mathcal{H}_1$ if

$$\begin{aligned}T_{\mathrm{naive}}(\mathbf{x})&=\frac{\mathbf{x}^{T}\hat{\mathbf{s}}_{\mathrm{naive}}}{\mathbf{x}^{T}\mathbf{x}}=\frac{\mathbf{x}^{T}\mathbf{H}(\hat{f}_{\mathrm{c,naive}})\hat{\boldsymbol{\theta}}_{\mathrm{naive}}}{\mathbf{x}^{T}\mathbf{x}}\\&\approx\frac{2\mathbf{x}^{T}\mathbf{H}(\hat{f}_{\mathrm{c,naive}})\mathbf{H}^{T}(\hat{f}_{\mathrm{c,naive}})\mathbf{x}}{N\mathbf{x}^{T}\mathbf{x}}>\eta_{\mathrm{naive}}.\end{aligned} \tag{34}$$

Note that the denominator $\mathbf{x}^{T}\mathbf{x}$ in (32) and (34) is in fact the estimation of noise energy under $\mathcal{H}_0$, and it serves as a normalization term. Also note that the test statistics of the two detectors have a similar form to the normalized periodogram detector proposed in [30] under unknown WGN condition. Closed-form solutions to the test statistic distributions are not easy to derive at the moment, and more about the performances of the two detectors will be provided in Section IV.

C. Time-Frequency Domain Detector

In this subsection, we propose a practical TF domain ENF detector, termed as TF detector, which exploits more *a priori* knowledge of the ENF signal. Instead of considering the ENF as a constant across the whole recording, we assume that the ENF is approximately a constant only within a given short time interval. Such an interval determines the size of each overlapping processing frame. Let the frame size be N_{F}, frame step-size be Δ, assuming the number of frames is L and N is appropriately zero-padded ($L<N$), then we have

$$N_{\mathrm{F}}+(L-1)\Delta=N, \tag{35}$$

and we could denote the lth frame of the ENF signal by

$$s_{l}[n]=A[l]\cos(2\pi Tf[l]n+\phi[l]), \tag{36}$$

where $l\in\{0,1,\ldots,L-1\}$ and $n\in\{0,1,\ldots,N_{\mathrm{F}}-1\}$. Fig. 3 shows how (35) and (36) are obtained. For each frame of $x[n]$, denoted by $x_l[n]$, we obtain the corresponding dominant frequency using the LS solution (31), i.e.,

$$\hat{f}_{x}[l]=\arg\max_{f}\left|\sum_{n=0}^{N_{\mathrm{F}}-1}x_{l}[n]\exp\left(j2\pi Tfn\right)\right|^{2}, \tag{37}$$

which is fortunately independent of the amplitude and phase. Concatenating $\hat{f}_{x}[l]$ for each l, we then obtain an ENF-like TF curve. Denote such a TF peak calculation process by

a mapping $\mathcal{T}\{\cdot\}: \mathbb{R}_+^{N\times 1} \to \mathbb{R}_+^{L\times 1}$, where $\mathbb{R}_+$ is the positive real number set, then the ENF detection problem (2) becomes

$$\begin{aligned}\mathcal{H}_0 &: \hat{f}_x[l] = \mathcal{T}\{v[n]\},\\ \mathcal{H}_1 &: \hat{f}_x[l] = \mathcal{T}\{s[n] + v[n]\},\end{aligned} \tag{38}$$

and the distributions of $\hat{f}_x[l]$ under both hypotheses are of our interest, which will reveal the clues for a sound ENF detector.

Under $\mathcal{H}_0$, the mapping $\mathcal{T}$ merely calculates the dominant frequency component of each noise segment in $v[n]$. Note that ideally, the PSD of $v[n]$ is tailored from the flat PSD of the white noise $\tilde{v}[n]$, which is thus also flat but only within the passband of the bandpass filter. Recall that the width of the ideal passband is B Hz, it then follows that

$$\hat{f}_x[l] \sim \mathcal{U}\left[50 - \frac{1}{2}B, 50 + \frac{1}{2}B\right], \quad \forall l, \text{under } \mathcal{H}_0, \tag{39}$$

and

$$\begin{cases} E\{\hat{f}_x[l]; \mathcal{H}_0\} = 50 \text{ Hz},\\ \text{var}\{\hat{f}_x[l]; \mathcal{H}_0\} = \dfrac{1}{12}B^2 \text{ Hz}^2, \end{cases} \tag{40}$$

thus $\hat{f}_x[l]$ under $\mathcal{H}_0$ could be modeled as a WSS piecewise uniformly distributed process. Assuming ergodicity, we could approximate the expectations in (40) by the corresponding averages over time index l.

Under $\mathcal{H}_1$, the estimation of the dominant frequency is the estimation of a single tone from noisy observations. The variance of the unbiased estimator (37) is lower bounded by the Cramer-Rao lower bound (CRLB) under colored noise model. The derivation of the exact CRLB is complicated and not easy to be numerically obtained as shown in [29] and [31]. Instead, we use the asymptotic CRLB derived in [32],

$$\begin{aligned}\text{var}\{\hat{f}_x[l]; \mathcal{H}_1\} &\geq \text{CRLB}_f\\ &= \frac{12 \times 2\sigma_{\tilde{v}}^2 |H(f_x[l])|^2}{N_{\text{F}}^3 A[l]^2} \times \frac{1}{(2\pi T)^2}\\ &= \frac{12}{N_{\text{F}}^3 \text{SNR}_l} \times \frac{1}{(2\pi T)^2},\end{aligned} \tag{41}$$

where $|H(f_x[l])|^2 = 1$ for an ideal bandpass filter, and SNR_l is the local SNR, equivalent to (4) calculated within the lth processing frame without taking the logarithm. It can also be seen from (41) that the CRLB under the specific colored WSS noise is equal to one obtained from white noise case [28]. Therefore, we write the statistics of the frequency estimate under $\mathcal{H}_1$ as

$$\begin{cases} E\{\hat{f}_x[l]; \mathcal{H}_1\} = f[l] \text{ Hz},\\ \text{var}\{\hat{f}_x[l]; \mathcal{H}_1\} \geq \text{CRLB}_f \text{ Hz}^2, \end{cases} \tag{42}$$

for the lth frame. While the above processing treats $f[l]$ as an unknown constant within the lth frame, the whole ENF signal extracted from the recording exhibits a WSS Gaussian AR process, which has been discussed in [14] and [25]. Fortunately, the reference ENF signal can be directly recorded from an arbitrary power supply using an electric transformer, and the extraction and statistical analysis of the reference ENF yield its *a priori* knowledge, namely, the mean and variance of the ENF from a specific power grid, denoted by μ_{ENF} and σ_{ENF}^2 respectively, are available to us. Based on the recording of the reference ENF signal of central China grid, we have $\mu_{\text{ENF}} = \hat{f}_{\text{c,naive}} = 50$ Hz, and $\sigma_{\text{ENF}}^2 = 4.56 \times 10^{-4}$ Hz^2. Assuming ergodicity, we have the following noise-free approximations

$$\begin{cases} \dfrac{1}{L}\displaystyle\sum_{l=0}^{L-1} f[l] \approx \mu_{\text{ENF}} = 50,\\ \dfrac{1}{L-1}\displaystyle\sum_{l=0}^{L-1}\left(f[l] - \dfrac{1}{L}\sum_{l=0}^{L-1} f[l]\right)^2 \approx \sigma_{\text{ENF}}^2. \end{cases} \tag{43}$$

Combining (42) and (43), we could consider $\hat{f}_x[l]$ under $\mathcal{H}_1$ as the estimation of the WSS Gaussian distributed ENF $f[l]$, and for each frame indexed by l, there exists another Gaussian perturbation with zero mean and variance greater than CRLB_f. The variance of $\hat{f}_x[l]$ evaluated via the consideration of both random noise and random ENF is given by

$$\text{var}_f\left\{\text{var}_v\{\hat{f}_x[l]; \mathcal{H}_1\}\right\} \geq \text{CRLB}_f + \sigma_{\text{ENF}}^2. \tag{44}$$

Exploiting the slow-varying nature of the ENF and the fast-varying nature of noise IF, we expect that $\hat{f}_x[l]$ fluctuates more under $\mathcal{H}_0$ than under $\mathcal{H}_1$. Therefore, the TF domain test statistic could be set as the sample variance of the estimated frequency peaks, and the detector decides $\mathcal{H}_1$ if

$$T_{\text{TF}}(\mathbf{x}) = \frac{1}{L-1}\sum_{l=0}^{L-1}\left(\hat{f}_x[l] - \frac{1}{L}\sum_{l=0}^{L-1}\hat{f}_x[l]\right)^2 < \eta_{\text{TF}}, \tag{45}$$

where $\hat{f}_x[l]$ is obtained from (37). Note that for the TF detector, the test statistics under $\mathcal{H}_0$ are generally greater than those under $\mathcal{H}_1$ thus the "$<$" sign is used in (45). For the ease of mathematical tractability, we approximate (39) by

$$\hat{f}_x[l] \sim \mathcal{N}\left(50, (B/6)^2\right), \quad \forall l, \text{under } \mathcal{H}_0, \tag{46}$$

so that the region $[50 - B/2, 50 + B/2]$ contains the mean plus and minus three times of the standard deviation, then the test statistic $T_{\text{TF}}(\mathbf{x})$ under $\mathcal{H}_0$ follows a chi-square distribution

$$\frac{(L-1)T_{\text{TF}}(\mathbf{x})}{(B/6)^2} \sim \chi_{L-1}^2, \text{ under } \mathcal{H}_0. \tag{47}$$

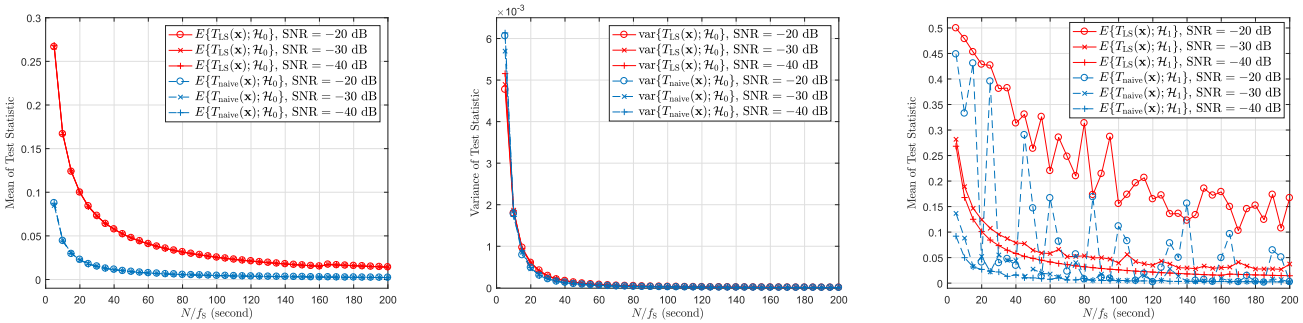
In the meantime, according to (42)–(44), we have

$$\hat{f}_x[l] \sim \mathcal{N}\left(50, \text{CRLB}_f + \sigma_{\text{ENF}}^2\right), \quad \forall l, \text{under } \mathcal{H}_1, \tag{48}$$

where the inequality in (44) is implicitly omitted assuming a minimum variance unbiased estimator (MVUE) is attained, thus the test statistic $T_{\text{TF}}(\mathbf{x})$ under $\mathcal{H}_1$ follows another chi-square distribution

$$\frac{(L-1)T_{\text{TF}}(\mathbf{x})}{\text{CRLB}_f + \sigma_{\text{ENF}}^2} \sim \chi_{L-1}^2, \text{ under } \mathcal{H}_1. \tag{49}$$

According to the theorem of functions of random variables [27] and after some manipulations, $T_{\text{TF}}(\mathbf{x})$ follows the



(a) Mean, $\mathcal{H}_0$. (b) Variance, $\mathcal{H}_0$. (c) Mean, $\mathcal{H}_1$.

Fig. 4. Empirical results of the mean and variance values of $T_{\mathrm{LS}}(\mathbf{x})$ and $T_{\mathrm{naive}}(\mathbf{x})$ versus data recording length under $\mathcal{H}_0$ and $\mathcal{H}_1$, where $f_{\mathrm{S}} = 400$ Hz, averaged from 10000 independent trials. (a) Mean values of $T_{\mathrm{LS}}(\mathbf{x})$ and $T_{\mathrm{naive}}(\mathbf{x})$ under $\mathcal{H}_0$; (b) Variance values of $T_{\mathrm{LS}}(\mathbf{x})$ and $T_{\mathrm{naive}}(\mathbf{x})$ under $\mathcal{H}_0$; (c) Mean values of $T_{\mathrm{LS}}(\mathbf{x})$ and $T_{\mathrm{naive}}(\mathbf{x})$ under $\mathcal{H}_1$. For brevity, the variance values under $\mathcal{H}_1$ are not presented.

following gamma distributions

$$T_{\mathrm{TF}}(\mathbf{x}) \sim \begin{cases} \mathcal{G}\left(\dfrac{L-1}{2}, \dfrac{2(B/6)^2}{L-1}\right), & \text{under } \mathcal{H}_0, \\ \mathcal{G}\left(\dfrac{L-1}{2}, \dfrac{2\left(\mathrm{CRLB}_f + \sigma^2_{\mathrm{ENF}}\right)}{L-1}\right), & \text{under } \mathcal{H}_1, \end{cases} \tag{50}$$

and we further have

$$\begin{cases} E\left\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_0\right\} \approx (B/6)^2, \\ E\left\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_1\right\} \geq \mathrm{CRLB}_f + \sigma^2_{\mathrm{ENF}}, \\ \mathrm{var}\left\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_0\right\} \approx \dfrac{2(B/6)^4}{L-1}, \\ \mathrm{var}\left\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_1\right\} \geq \dfrac{2\left(\mathrm{CRLB}_f + \sigma^2_{\mathrm{ENF}}\right)^2}{L-1}. \end{cases} \tag{51}$$

According to (35), the degrees of freedom are calculated by $L - 1 = (N - N_{\mathrm{F}})/\Delta$, which is usually a relatively large quantity. With large degrees of freedom, T_{TF} approximates Gaussian distributions under both hypotheses. The equalities in (51) hold for an MVUE. We analyze the performances of the derived detectors in the next section.

IV. Performance Analysis

A. Deciding the Thresholds

To identify the optimal threshold values, we analyze and compare the test statistics distributions of the practical detectors derived in Section III. Because both signal and noise parameters are unknown, we are interested in finding test statistics independent of the unknown parameters under $\mathcal{H}_0$ and then a CFAR detector could be obtained.

1) Thresholds for LS-LRT and Naive-LRT: Since $T_{\mathrm{LS}}(\mathbf{x})$ and $T_{\mathrm{naive}}(\mathbf{x})$ have the same form except for the estimated frequency, we focus on $T_{\mathrm{LS}}(\mathbf{x})$ for the moment. Substituting the original form of $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ in (30) into (32), we have

$$T_{\mathrm{LS}}(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\mathbf{x}^T\mathbf{P}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\|\mathbf{P}\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}, \tag{52}$$

where $\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ is the projection matrix to the column space of $\mathbf{H}$, and $\mathbf{P}^T\mathbf{P} = \mathbf{P}$. Therefore, the test statistic of the LS-LRT detector is in fact a normalized projector which projects the observed data $\mathbf{x}$ on its strongest frequency component obtained by (31), while the test statistic of the naive-LRT detector projects $\mathbf{x}$ on the nominal frequency 50 Hz. Under $\mathcal{H}_0$, recall that the noise $\upsilon[n]$ has a band limited but flat PSD within the passband of the bandpass filter $h[n]$. The $\hat{f}_{\mathrm{c,LS}}$ obtained by (31) is hence a random number within the passband B and it is expected that (32) or equivalently (52) yields a very small fraction. In other words, the energy of the component of $\mathbf{x}$ lying in the column space of $\mathbf{H}$ is expected to be a very small portion of the total energy of $\mathbf{x}$. Further, the ratio $\|\mathbf{P}\mathbf{x}\|_2^2/\|\mathbf{x}\|_2^2$ tends to decrease when N is increased because of space expansion.

The above analysis is evidenced by the simulation results presented in Fig. 4. In this figure, it is observed that first, the test statistics for both LS-LRT and naive-LRT detectors are independent of SNR values, and second, the mean and variance values under $\mathcal{H}_0$ are a monotonically decreasing function of data recording time. In contrast, under $\mathcal{H}_1$, the mean values of $T_{\mathrm{LS}}(\mathbf{x})$ and $T_{\mathrm{naive}}(\mathbf{x})$ become dependent of the unknown SNR value, as shown in Fig. 4 (c). The variance curves under $\mathcal{H}_1$ are not presented for brevity. Because the naive-LRT simply uses the nominal value as the frequency estimate, the resultant energies of the projections are smaller than those with the MLE strongest frequency component in the LS-LRT, and this could be observed in Fig. 4 (a). In addition, It could be observed from Fig. 4 (c) that under a higher SNR, the $\mathcal{H}_1$ $T_{\mathrm{LS}}(\mathbf{x})$ mean curve tends to be separated further away from the $\mathcal{H}_0$ $T_{\mathrm{LS}}(\mathbf{x})$ mean curve in Fig. 4 (a), thus allowing better detection performance. However, for the naive-LRT detector the $\mathcal{H}_1$ statistic is unstable.

Although the test statistic distributions under $\mathcal{H}_0$ are unknown, the corresponding means and variances are invariant to the SNR and are functions of N. This allows us to establish practical thresholds for ENF detection. We now denote the thresholds for LS-LRT and naive-LRT as a function of N and according to the empirical results in Fig. 4 (a), (b), set

$$\eta_{\mathrm{LS}}(N) = E\{T_{\mathrm{LS}}(\mathbf{x}); \mathcal{H}_0\} + \alpha\sqrt{\mathrm{var}\{T_{\mathrm{LS}}(\mathbf{x}); \mathcal{H}_0\}}, \tag{53}$$

$$\eta_{\mathrm{naive}}(N) = E\{T_{\mathrm{naive}}(\mathbf{x}); \mathcal{H}_0\} + \alpha\sqrt{\mathrm{var}\{T_{\mathrm{naive}}(\mathbf{x}); \mathcal{H}_0\}}, \tag{54}$$

where $\alpha > 0$ is a weight and N is the length of $\mathbf{x}$. The simulated accuracy values (number of correct detections over

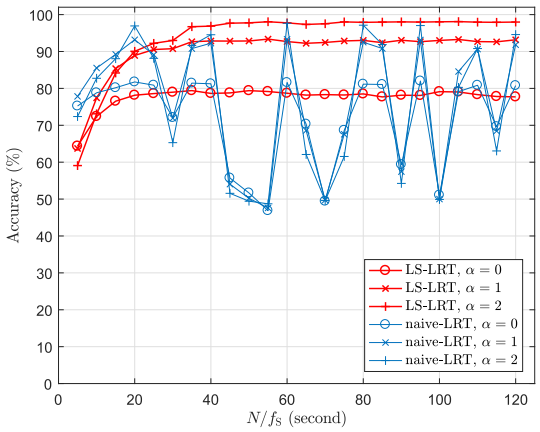


Fig. 5. Detection accuracy versus data recording time for LS-LRT and naive-LRT detectors, averaged by 10000 independent trials, SNR $= -25$ dB.

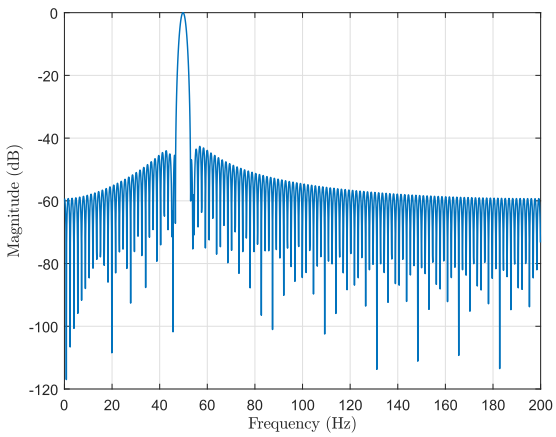


Fig. 6. Magnitude response of the linear phase FIR bandpass filter $h[n]$, where $M = 255$, the approximate passband is $2 < B < 3$ Hz, and it follows from (51) that $0.11 < E\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_0\} \approx (B/6)^2 < 0.25$ Hz under $\mathcal{H}_0$.

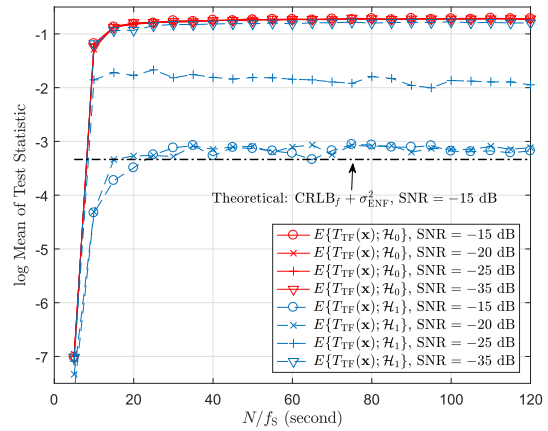
total number of trials) versus data recording length using different values of α are shown in Fig. 5. It could be seen that $\alpha = 2$ is a sound choice for the LS-LRT threshold. However, unstable detection performance of naive-LRT is observed, but it is specially noticed that for very short recordings, e.g., duration less than 10 seconds in Fig. 5, naive-LRT outperforms LS-LRT in terms of detection accuracy when $\alpha = 1$ or 2. The reason behind mainly lies in the limitation of the frequency resolution. For short recordings, the observation time (in seconds) is insufficient for the Fourier transform to resolve the true ENF. Thus the use of the nominal value in naive-LRT, although intuitive and straightforward, could reduce the frequency estimation error for very short recordings. Another fact that may explain this performance is the slowly varying nature of the ENF, i.e., for short recordings, the ENF tends to be a constant. More results on ENF detection in very short recordings are presented in Section IV-B.

2) Thresholds for the TF Detector: For the TF detector, the threshold value could be theoretically approximated by examining (51). The test statistic (45) is the sample variance of the signal formed by concatenated frequency peaks of the observed data frames. Under $\mathcal{H}_0$, the observation is the bandpass filtered noise only, and it is important to notice that TF domain test statistic becomes invariant to the unknown noise power. In fact, the mean value of $T_{\mathrm{TF}}(\mathbf{x})$, i.e., $E\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_0\}$, is fully characterized by the frequency response of the bandpass filter $h[n]$, while the variance $\mathrm{var}\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_0\}$ is determined by $h[n]$ and the number of frames L. The magnitude response of the linear phase FIR bandpass filter $h[n]$ for the practical detectors is depicted in Fig. 6. Instead of being ideally flat in its passband, the shape is better approximated by a Gaussian distribution in (46). Under $\mathcal{H}_1$, the test statistic is dependent on noise parameter. Based on the WSS signal model and for mathematical tractability, we replace the local SNR, i.e., SNR_l by the global SNR in (4). For a fixed STFT stepsize Δ and frame size N_{F}, we rewrite (35) as

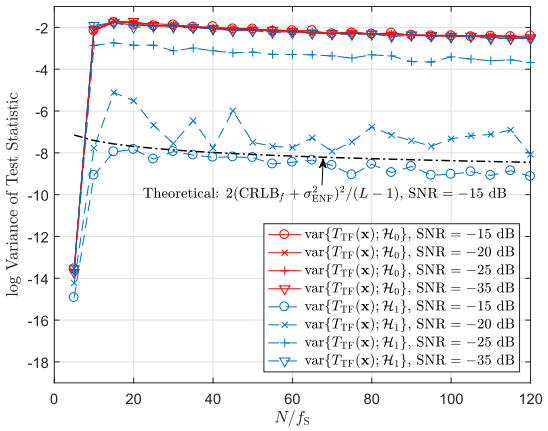
$$L = \frac{N - N_F}{\Delta} + 1, \tag{55}$$

thus L is a liner function of N. Given this bandpass filter $h[n]$ and the number of frames L, the best detection situation for the TF detector could be quantified by setting $\mathrm{CRLB}_f \to 0$ in (51) if $\mathrm{SNR} \to \infty$ or N_{F} is sufficiently large (see (41)), or if $\mathrm{CRLB}_f \ll \sigma^2_{\mathrm{ENF}}$ and could be discarded.

Fig. 7 depicts the simulation results of the test statistic distributions of the TF detector versus data recording length under different SNR values, which are consistent with the theoretical results in (51) in the following aspects. **i)** The TF detector is not suitable for very short recordings with durations less than 20 seconds, which is interestingly opposite to case of the naive-LRT detector shown in Fig. 5. When the data recording length becomes greater than 20 seconds, the asymptotic property is satisfied and the performance converges. **ii)** The mean values of $T_{\mathrm{TF}(\mathbf{x})}$ are independent of N under both hypotheses, which is observed in Fig. 7 (a). The mean values under $\mathcal{H}_0$ are further independent of the SNR values as the (red) mean curves overlap at a level of about $0.2 \approx 10^{-0.7}$, which is consistent with the theoretical range $0.11 < E\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_0\} \approx (B/6)^2 \approx 0.2 < 0.25$. **iii)** The mean value under $\mathcal{H}_1$ is dependent on the SNR value and it approaches the $\mathcal{H}_0$ level as SNR decreases. In addition, we observe that the mean curves for SNR $= -15$ and -20 dB are at the same level. This is because of the term σ^2_{ENF} does not allow the mean curve to further decrease. The theoretical mean curve when SNR $= -15$ dB is also presented in Fig. 7 (a). Therefore, given $h[n]$ and N_{F}, the best possible separation of the means of $T_{\mathrm{TF}(\mathbf{x})}$ between the two hypotheses is the distance between the red curve level and the theoretical bound in Fig. 7 (a). **iv)** Similarly, consistent results are observed in Fig. 7 (b) for the test statistic variances which are inversely proportional to $L - 1$. Note that the curve for SNR $= -15$ dB under $\mathcal{H}_1$ is lower than the theoretical values. This is because (51) is based on the assumed Gaussian distributed approximation of $\hat{f}_x[l]$ which is strictly not Gaussian. For example, $\hat{f}_x[l]$ does not take any value in the stopband of $h[n]$, thus the assumed Gaussian distribution tails are truncated, resulting in reduced variances. **v)** Finally, when SNR < -20 dB, the STFT-based frequency estimator suffers from the threshold effect [33], resulting in that the simulated test statistic mean and variance curves become much greater than the



(a) Mean, $\mathcal{H}_0$ and $\mathcal{H}_1$.



(b) Variance, $\mathcal{H}_0$ and $\mathcal{H}_1$.

Fig. 7. Empirical results of log mean and variance values of $T_{\mathrm{TF}}(\mathbf{x})$ versus data recording length under $\mathcal{H}_0$ and $\mathcal{H}_1$, where $f_S = 400$ Hz, $\Delta = f_S$, $N_F = 16 f_S$, averaged from 1000 independent trials. (a) log mean values of $T_{\mathrm{TF}}(\mathbf{x})$ under both hypotheses; (b) log variance values of $T_{\mathrm{TF}}(\mathbf{x})$ under both hypotheses. The simulated variances when SNR $= -15$ dB under $\mathcal{H}_1$ are less than the theoretical values because $\hat{f}_x[l]$ is not strictly Gaussian distributed.

CRLB levels. Therefore, theoretical curves for low SNR values are not provided.

Based on the above analysis, statistical information about $T_{\mathrm{TF}}(\mathbf{x})$ under $\mathcal{H}_0$ approximately follows a gamma distribution with all parameters known, and with a sufficiently large L, it approaches a Gaussian distribution, thus the TF detector is a CFAR detector. Generally, we set the TF threshold as follows

$$\begin{aligned}\eta_{\mathrm{TF}}(L) &= E\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_0\} + \beta\sqrt{\mathrm{var}\{T_{\mathrm{TF}}(\mathbf{x}); \mathcal{H}_0\}} \\ &\approx (B/6)^2 - \beta\sqrt{\frac{2(B/6)^4}{L-1}}, \end{aligned} \tag{56}$$

where $\beta > 0$ is a weight similar to α. Note that the TF threshold is a function of L. It is also dependent on the magnitude response of $h[n]$ which is parameterized using B and treated as a fixed term here. If a different bandpass filter is used, then $\eta_{\mathrm{TF}}(L)$ could be computed accordingly.

B. In the Presence of Content Tampering

Although ENF detection and ENF-based audio tampering detection are two very different problems and the latter is not in the scope of this paper, we are interested to see how ENF detection would be affected when the recording has been tampered with. We first consider deletion attack and then briefly discuss about other types of attacks.

i) For deletion attack. On the one hand, for the LS-LRT and naive-LRT detectors, the test statistics (32) and (34) are in the form of normalized energy (see the denominators therein). These test statistics are hence invariant to deletion attack. This could be verified from Fig. 4 as we could see that as the observation time N/f_S increases (asymptotic property satisfied), the values of test statistics under both hypotheses tend to converge respectively. In this situation, deletion attack has the effect of reducing the value of N/f_S. As long as N/f_S after deletion still lies in the "flat" region, the detection performance is unaffected. On the other hand, the proposed TF domain detector uses an estimated TF domain variance (45) as the test statistic. Under both hypotheses, it is known that both noise and ENF are stationary, thus the effect of deletion attack on the TF domain detector is in fact the removal of a portion of the observed stationary process, which statistically does not alter the mean and variance. Therefore, the performance of the TF detector is also invariant to deletion attack, as can be observed from Fig. 7 that the test statistics are nearly flat given sufficient observation times.

ii) For other types of attacks such as insertion and replacement, the substituted signal segments are out of our control. If the adversary applied anti-forensic manipulation by inserting an ENF segment together with the substituted audio content, then ENF detection could still be unaffected thanks to statistical consistency, and the problem becomes after successful ENF detection, how to effectively detect the manipulated ENF segment, as discussed in [34]. However, if the substituted signal segments do not contain ENF signal, then ENF detection performance will deteriorate. This problem is noted here for future research attention.

C. Performance Comparison

In this subsection, we compare the detection performances of the proposed ENF detectors using synthetic data. We first examine the detection accuracies of the proposed detectors using the theoretically derived decision thresholds, and the corresponding results are provided in Fig. 8 for different data recording lengths, where $\alpha = \beta = 2$ and SNR $= -25$ dB. It could be seen that the detection performance is much better when $s[n]$ is known (MF curve) than when $s[n]$ is unknown for short recordings. The LS-LRT detector achieves 95% accuracies when the recording duration is longer than 30 seconds, while the TF detector approaches the MF bound when the recording duration is longer than 80 seconds. In contrast, the naive-LRT is unstable in most situations.

To further illustrate the influence of the data recording length on detection performance, we examine the ROC curves of the proposed detectors for very short recordings of 5 seconds and long recordings of 300 seconds respectively, under different SNR values, and the corresponding results are shown in Figs. 9 and 10 respectively. From Fig. 9, we observe consistent results with those shown in Figs. 5 and 8, that the naive-LRT yields better detection performance than other

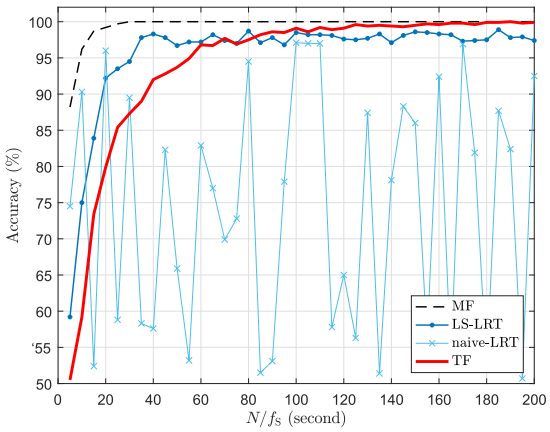


Fig. 8. Detection accuracy versus data recording length, where $f_S = 400$ Hz, $\Delta = f_S$, $N_F = 16f_S$, $\alpha = \beta = 2$, SNR $= -25$ dB, averaged from 1000 independent trials.

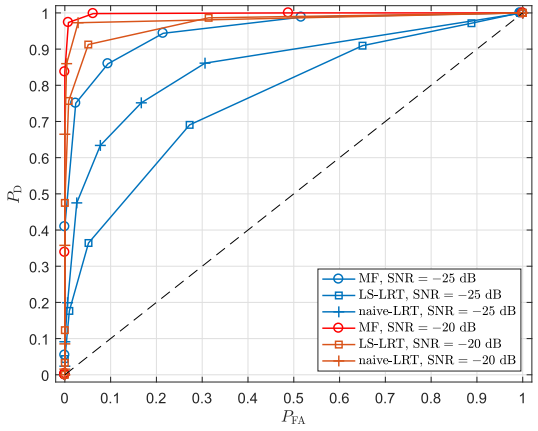


Fig. 9. ROC curves of LS-LRT and naive-LRT detectors for 5-second recordings under different SNR values, where $f_S = 400$ Hz, averaged from 1000 independent trials.

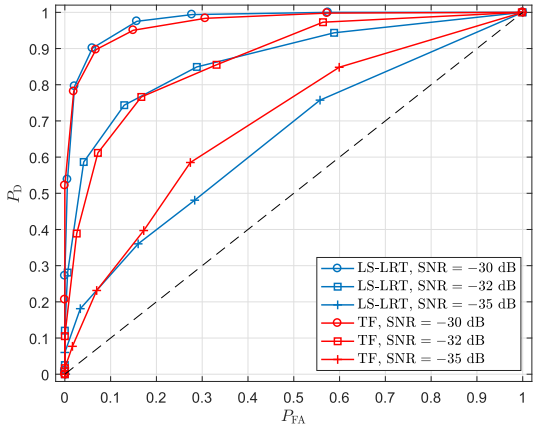


Fig. 10. ROC curves of LS-LRT and TF detectors for 300-second recordings under different SNR values, where $f_S = 400$ Hz, $\Delta = f_S$, $N_F = 16f_S$, $\alpha = \beta = 2$, averaged from 1000 independent trials.

competitors for 5-second recordings. It approaches the MF detection bound when SNR $= -20$ dB. In Fig. 10, the LS-LRT and TF detectors perform equally well when SNR $= -30$ dB. As the SNR value further decreases, both detectors suffer from

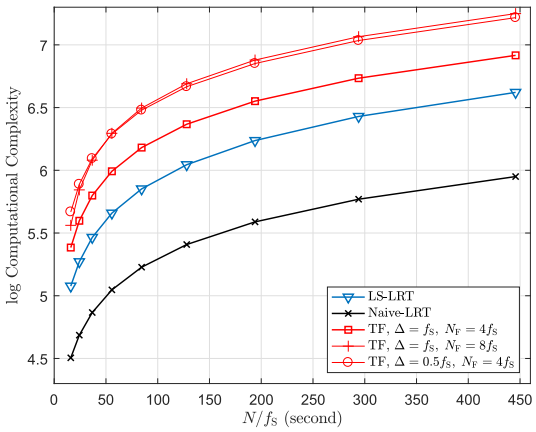


Fig. 11. Computational complexity vs. recording data length.

performance degradation. The ENF detection results of the LS-LRT detector indicate that a long data recording length N is highly advantageous in detecting sinusoidal signal in noise, and this is interestingly also true even for a sinusoidal with slightly varying IFs, i.e., the ENF. Recall (18), the LS-LRT detector assumes that the ENF is a constant throughout the observation time.

D. Computational Complexity

In this subsection, we briefly analyze the computational complexity of the detectors derived in this paper. Note that the complexity of the MF-like detector is derived from a known ENF signal, and it is not comparable with the practical detectors. Besides, the complexity of the TF-CRLB detector is dependent on the complexity of the MVUE and is hence not given. For the LS-LRT detector, the calculation of the periodogram in (31) has a complexity of $O(N \log_2 N)$ and the peak search has a complexity of $O(N)$. Besides, the calculation of $T_{\text{LS}}(\mathbf{x})$ in (32) yields a complexity of $O(5N + 2)$. Therefore, the computational complexity of the LS-LRT detector is

$$\text{Complexity}_{\text{LS-LRT}} = O(N \log_2 N + 6N + 2). \tag{57}$$

For the naive-LRT, the frequency estimation procedure is omitted, and the computational complexity only comes from the calculation of $T_{\text{naive}}(\mathbf{x})$ in (34), which is

$$\text{Complexity}_{\text{naive-LRT}} = O(5N + 2). \tag{58}$$

For the TF detector, the computational complexity depends on the number of frames L and frame size N_F, and the sample variance (45) has a complexity of L. Thus we have

$$\text{Complexity}_{\text{TF}} = O\left(L\left(N_F \log_2 N_F + N_F + 1\right)\right). \tag{59}$$

The comparison results of computational complexity as a function of N are shown in Fig. 11 where the complexity values are scaled by $\log_{10}(\cdot)$. It can be seen that the computational complexity of the TF detector is several times higher than that of the LS-LRT detector across N when the frame step-size is 1 second, i.e., $\Delta = f_S$, and the frame size is 4 seconds, i.e., $N_F = 4f_S$. If the step-size is reduced to half

TABLE I
SUMMARY OF DETECTORS

	Detector	A, ϕ Estimator	f Estimator	Test Statistic	Threshold	Frame	CFAR	Complexity
Bounds	MF-Like	Known	Known	(9)	$\eta_{\text{MF}} = 1/2\|\mathbf{s}\|_2^2$	Single	No	$O(N)$
	TF-CRLB	Unnecessary	$\hat{f}_{\text{MVUE}}$	$(45)\|_{\hat{f}_x=\hat{f}_{\text{MVUE}}}$	η_{TF}: (56)	Multiple	Yes	–
Practical	LS-LRT	(30)	(31)	(32)	η_{LS}: (53)	Single	No	(57)
	Naive-LRT	(30)	$\hat{f}_{\text{c,naive}} = 50$	(34)	η_{naive}: (54)	Single	No	(58)
	TF	Unnecessary	(37)	(45)	η_{TF}: (56)	Multiple	Yes	(59)

Note: The most efficient calculation of the numerator of (32) and (34) is $(\mathbf{x}^T\mathbf{H})(\mathbf{H}^T\mathbf{x})$, whose computational complexity is $O(4N+2)$.

a second or the frame size is doubled, then the computational complexity of the TF detector nearly doubles. The reason for the TF detector to have increased computational loads mainly lies in the frame-based processing. Note that the LS-LRT and TF detectors yield very competitive detection results, while the former is more computationally efficient and the latter is approximately a CFAR detector. The details and properties of the proposed detectors are summarized in Table I.

V. Experimental Results

In this section, we evaluate the performances of the proposed ENF detectors using a dataset of 60 real-world audio recordings made around Wuhan University campus, sampled at 44100 Hz with 16-bit quantization and mono channel.[2] The recording locations include classrooms, meeting rooms, graduate student offices, campus paths, main streets, dormitories, libraries, etc., the recording duration spans 5 to 20 minutes, and the recording environments include both sunny and rainy days and nights. The recordings are made by voice recorders and smartphones, and we have found that as long as the recording program does not cut off frequencies below 100 Hz, differences among recording devices will not have noticeable effects on the quality of recorded data. Among the 60 recordings, 50 have ENF captured and verified according to the known recording times and comparison with the corresponding reference database. The other 10 are usually made in wide-open exterior environments (strong noise and interference) and the user is walking (Doppler effects), and we have verified against the reference database that they contain severely corrupted or no ENF signal. Therefore, the ground truth information about ENF existence in these recordings is known to us, which could be used to evaluate the performances of different ENF detectors.

The recordings are cropped at random locations to construct a larger dataset with duration values varying from 5 to 270 seconds, and for each duration, we randomly crop 50 clips for $\mathcal{H}_0$ and 50 clips for $\mathcal{H}_1$ respectively. For better noise control, the length of bandpass filter is increased from 256 to 1024, with a narrower passband and further attenuated stopband compared with the one in Fig. 6. The performances

[2]The dataset and MATLAB programs of this paper could be downloaded from https://github.com/ghuawhu/ENF-WHU-Dataset.

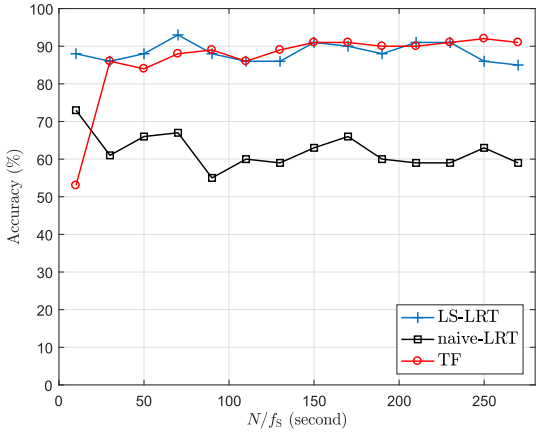


Fig. 12. Detection accuracy versus data recording length using real audio recordings, where $f_S = 400$ Hz, $\alpha = \beta = 2$, $\Delta = f_S$, and $N_F = 16 f_S$.

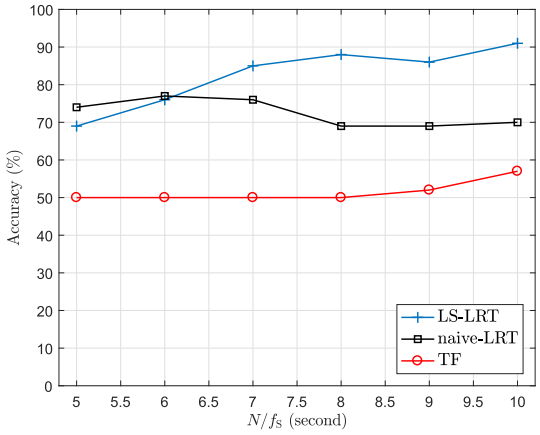


Fig. 13. Detection accuracy for short real audio recordings, where $f_S = 400$ Hz, $\alpha = \beta = 2$, $\Delta = f_S$, and $N_F = 16 f_S$.

of the proposed detectors are first measured by detection accuracy, which are depicted in Figs. 12 and 13 respectively. The detection accuracy is defined by the ratio of the number of correct detection results to the total number of recordings. Because the to-be-detected ENF signal is unknown in practical situation, the MF-like detector could not be implemented. It could be seen from the two figures that the performances of the proposed three detectors are generally consistent with the synthetic results shown in Figs. 8-10. The LS-LRT and TF

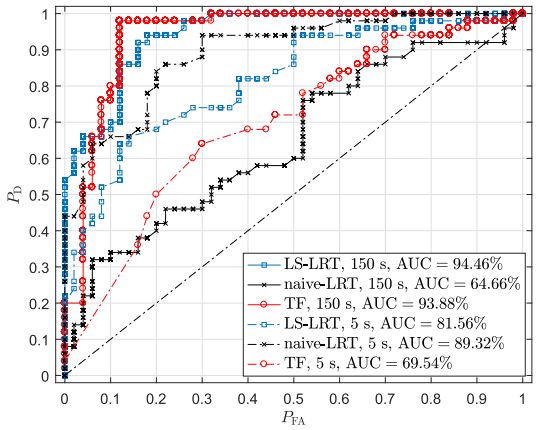


Fig. 14. ROC curves and AUC values of LS-LRT, naive-LRT, and TF detectors using real-world audio recordings, where duration values are 150 (solid) and 5 (dashed) seconds respectively, $\Delta = f_{\mathrm{S}}$, and $N_{\mathrm{F}} = 16\ f_{\mathrm{S}}$.

detectors achieve very competitive detection results for recordings longer than 50 seconds, while the naive-LRT detector only yields results slightly better than random guesses. For very short recordings, the naive-LRT detector yields the best accuracy, but only for 5-second and shorter recordings.

Further, applying random sampling, we constructed another two datasets consisting of 100 clips, having 150 seconds and 5 seconds duration values, 50 under $\mathcal{H}_0$ and 50 under $\mathcal{H}_1$ respectively. The results measured by ROC and area under ROC curve (AUC) values are shown in Fig. 14. For 150-second recordings, the TF detector yields consistently the best performance under varying threshold values, while LS-LRT yields slightly lowered performance. In contrast, for 5-second recordings, the TF detector turns out to be the worst performer due to the lack of observation time. Instead, the naive-LRT detector achieves the best trade-off between false alarms and true positive decisions.

VI. Conclusion

Although the ENF has been found to be an effective forensic criterion to authenticate digital multimedia files, its existence in these files has not been well addressed, especially for the most applicable format of audio. In this paper, we have comprehensively studied the problem of ENF detection in audio recordings, from theoretical perspective to practical detectors. We first assumed that the ENF signal is known and derived the MF-like detector as the clairvoyant detector over the GMF and its asymptotic approximation. Then, we approximated the model (3) by assuming the unknown ENF parameters as constants in (18) instead of random processes, and derived two LS-based detectors, i.e., LS-LRT (32) and naive-LRT (34) respectively. In addition, assuming the ENF is constant within a short time interval in (36), we proposed a frame-based TF domain detector in (45) exploiting the *a priori* knowledge of the ENF, and the performance reaches its best if the ENF estimator is an MVUE. Based on the derived test statistic distributions, the decision thresholds are given in (53), (54), and (56), respectively. Both synthetic and practical experiments have been carried out for performance evaluation, and it is found that the LS-LRT and TF detectors yield very competitive detection results for normal and long recordings, while the naive-LRT detector only outperforms the other detectors for very short recordings. For very short recordings, the use of the nominal value as the estimate of the ENF has been shown to be a good remedy for ENF detection.

In ENF-based forensic analysis, the duration of a questioned recording is usually preferred to be long (e.g., 10 minutes and longer in works including [2], [6], [16], [18]–[21], [35], [36]) to ensure better ENF extraction. This highly limits the ENF applicability as the duration of an audio evidence could indeed be short. In this paper, we have shown that it is possible to detect ENF in relatively short recordings (e.g., less than 2 minutes), and this indicates the possibility to perform reliable ENF estimation and further forensic analysis for short recordings. Towards this direction, high resolution (instantaneous) frequency estimators may be incorporated to replace the Fourier analysis, which could potentially improve both ENF detection and estimation.

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] 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 Security*, vol. 8, no. 9, pp. 1417–1432, Sep. 2013.

[3] 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.

[4] G. Hua, J. Goh, and V. L. L. Thing, "A dynamic matching algorithm for audio timestamp identification using the ENF criterion," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 7, pp. 1045–1055, Jul. 2014.

[5] G. Hua, "Error analysis of forensic ENF matching," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2018, pp. 1–7.

[6] L. Zheng, Y. Zhang, C. E. Lee, and V. L. L. Thing, "Time-of-recording estimation for audio recordings," in *Proc. Digit. Forensic Res. Conf. (DFRWS US)*, 2017, pp. 1–11.

[7] D. P. Nicolalde 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.

[8] J. Chai, Y. Liu, Z. Yuan, R. W. Conners, and Y. Liu, *Tampering Detection of Digital Recordings Using Electric Network Frequency and Phase Angle*, vol. 135. New York, NY, USA: Audio Engineering Society, Inc., Oct. 2013. [Online]. Available: https://secure.aes.org/forum/pubs/conventions/?elib=17046

[9] P. A. A. Esquef, J. A. Apolinario, and L. W. P. Biscainho, "Edit detection in speech recordings via instantaneous electric network frequency variations," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2314–2326, Dec. 2014.

[10] P. M. G. I. Reis, J. P. C. L. da Costa, R. K. Miranda, and G. Del Galdo, "ESPRIT-Hilbert-Based audio tampering detection with SVM classifier for forensic analysis via electrical network frequency," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 4, pp. 853–864, Apr. 2017.

[11] 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.

[12] Y. Hu, C.-T. Li, Z. Lv, and B.-B. Liu, "Audio forgery detection based on max offsets for cross correlation between ENF and reference signal," in *Proc. Int. Workshop Digit. Forensics Watermarking (IWDW)*, Oct. 2013, pp. 253–266.

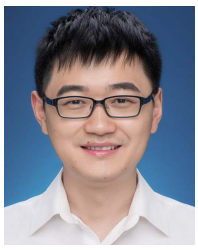
[13] A. Hajj-Ahmad, A. Berkovich, and M. Wu, "Exploiting power signatures for camera forensics," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 713–717, May 2016.

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

[15] W. Yao *et al.*, "Source location identification of distribution-level electric network frequency signals at multiple geographic scales," *IEEE Access*, vol. 5, pp. 11166–11175, 2017.

[16] 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.

[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] 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. 84:1–84:13, Nov. 2018.
[22] 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.
[23] 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.
[24] S. K. Mitra, *Digital Signal Processing A Computer-Based Approach*, 3rd ed. New York, NY, USA: McGraw-Hill, 2006.
[25] R. Garg, A. L. Varna, and M. Wu, "Modeling and analysis of electric network frequency signal for timestamp verification," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2012, pp. 67–72.
[26] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
[27] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill, 2002.
[28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[29] P. Stoica, A. Jakobsson, and J. Li, "Cisoid parameter estimation in the colored noise case: Asymptotic cramer-rao bound, maximum likelihood, and nonlinear least-squares," *IEEE Trans. Signal Process.*, vol. 45, no. 8, pp. 2048–2059, Aug. 1997.
[30] S. M. Kay and J. R. Gabriel, "Optimal invariant detection of a sinusoid with unknown parameters," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 27–40, Jan. 2002.
[31] J. M. Francos and B. Friedlander, "Bounds for estimation of complex exponentials in unknown colored noise," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2176–2185, Sep. 1995.
[32] P. Stoica and A. Nehorai, "Statistical analysis of two nonlinear least-squares estimators of sine-wave parameters in the colored-noise case," *Circuits, Syst., Signal Process.*, vol. 8, no. 1, pp. 3–15, Mar. 1989.
[33] 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.
[34] W.-H. Chuang, R. Garg, and M. Wu, "Anti-forensics and countermeasures of electrical network frequency analysis," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 12, pp. 2073–2086, Dec. 2013.
[35] 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.
[36] Y. Liu, Z. Yuan, P. N. Markham, R. W. Conners, and Y. Liu, "Wide-area frequency as a criterion for digital audio recording authentication," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2011, pp. 1–7.



Guang Hua (Member, IEEE) received the B.Eng. degree in communication engineering from Wuhan University, China, in 2009, and the M.Sc. degree in signal processing and the Ph.D. degree in information engineering from Nanyang Technological University, Singapore, in 2010 and 2014, respectively. From July 2013 to November 2015, he was a Research Scientist with the Department of Cyber Security and Intelligence, Institute for Infocomm Research, Singapore. After that, he was with the School of Electrical and Electronics Engineering, Nanyang Technological University, as a Research Fellow, until 2017. He is currently with the School of Electronic Information, Wuhan University. His research interests include multimedia forensics and security, applied convex optimization, applied machine learning, and general signal processing topics. He serves as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.

Han Liao received the M.Sc. degree in information and communication engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2018. She is currently pursuing the Ph.D. degree with the Department of Information and Communication Engineering, School of Electronic Information, Wuhan University, China. She has been working on robust electric network frequency (ENF) processing techniques for improved ENF-based multimedia forensics. Her research interests include signal detection, signal estimation, and multimedia tampering detection.

Qingyi Wang received the B.Sc. degree in antenna and propagation from Wuhan University, Wuhan, China, in 2019, where she is currently pursuing the M.Sc. degree in electronic and information engineering with the School of Electronic Information. Her research interests include multimedia security and forensics, and the software development of digital forensic systems. She has been working on the statistical analysis of electric network frequency (ENF) using signal processing and time series analysis tools. She was the recipient of the undergraduate scholarship from Wuhan University for four consecutive years.

Haijian Zhang (Member, IEEE) received the B.Sc. degree in electronic information engineering from Wuhan University, Wuhan, China, in 2006, and the joint Ph.D. degree from the Conservatoire National des Arts et Metiers, Paris, France, and Wuhan University, in 2011. From 2011 to 2014, he was a Research Fellow of the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore. He is currently an Associate Professor with the School of Electronic Information, Wuhan University. His main research interests include time-frequency analysis, array signal processing, and compressive sensing.

Dengpan Ye (Member, IEEE) received the B.Sc. degree in automatic control from SCUT in 1996 and the Ph.D. degree from NJUST in 2005. He was a Post-Doctoral Fellow of the School of Information System, Singapore Management University. Since 2012, he has been a Professor with the School of Cyber Science and Engineering, Wuhan University. He has authored or coauthored over 30 refereed journal articles and conference papers. His research interests include machine learning and multimedia security.