

Received November 14, 2016, accepted December 23, 2016, date of publication January 5, 2017, date of current version October 25, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2647921

On Practical Issues of Electric Network Frequency Based Audio Forensics

GUANG HUA¹, (Member, IEEE), GUOAN BI², (Senior Member, IEEE),
AND VRIZLYNN L. L. THING³, (Senior Member, IEEE)

¹School of Electronic Information, Wuhan University, Wuhan 430072, China

²School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798

³Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632

Corresponding author: Guoan Bi (egbi@ntu.edu.sg)

This work was supported in part by the Singapore Police Force, Ministry of Home Affairs, the Research under Grant CA/20150428/003 and in part by the Singapore Ministry of Education Research under Grant MOE2014-T2-1-079.

ABSTRACT The transmission frequency of power grids, i.e., electric network frequency (ENF), has become a common criterion to authenticate audio recordings during the past decade, drawing much attention from both the academic researchers and law enforcement agencies world widely. The properties of ENF enable forensic applications such as audio evidence timestamp verification and tampering detection. In this paper, based on a general review of existing works, we discuss several important practical problems and facts that have drawn less research attention or have not been formally studied, including ENF detection problems, limitations of the ENF-based tampering detection systems, and the difficulties in the ENF analysis. During ENF detection, the challenges come from not only the noise and the interference, but also the fact that audio recordings without captured ENF can still have signal components in the frequency band of interest (false positive). In ENF-based tampering detection systems, the weakness of commonly used assumptions and the limitations of several existing solutions are discussed. In addition, we reveal that in the most intensively studied ENF-based audio evidence timestamp verification, many works aiming at improving ENF estimation could only produce marginal performance improvement, while the main problems due to noise and interference remain open. All these analysis and discussions are related by a proposed big picture of ENF-based audio authentication systems. After that, we also investigate the strategies to design more reliable audio authentication systems based on the ENF, which consists of a series of research and investigation works.

INDEX TERMS Electric network frequency (ENF), audio forensics, audio timestamp verification, audio tampering detection, audio authentication.

I. INTRODUCTION

A. BACKGROUND

The advancements of digital and computer technologies have created today's highly connected world in which a huge and fast-increasing volume of multimedia content is flying through cyberspace to reach millions of end users. The ubiquity of digital data acquisition devices has brought new challenges to forensic analysts, which has lead to increasing research attention in the area of digital forensics. A typical example could be the need for the verification and authentication of evidences obtained in digital forms (image, audio, video, etc.) to ensure the corresponding admissibility during criminal investigation or other law enforcement cases.

In audio forensics, a sub-category of digital forensics, one of the recent most significant achievements would be the research and development of using the electric network frequency (ENF) as a kind of naturally embedded signature in many audio recordings for audio evidence authentication purposes [1]–[28]. ENF is the transmission frequency of power grids with a nominal value at either 50 or 60 Hz, and it has been observed to have two important properties, i.e., tiny random fluctuation and intra-grid consistency. The random fluctuation is caused by the varying load in the power grid, and it indicates that the ENF segments captured in audio recordings are unique. Making use of the intra-grid consistency, the ENF signal captured in a testing recording can be extracted and matched with a ground truth reference

obtained from an ENF database. This explains an application of audio timestamp verification based on the ENF criterion, while tampering detection and other applications are also possible to be established. Compared to other existing audio forensic solutions, e.g., compression detection [29], reverberation measurement [30], microphone classification [31], etc., ENF based solutions possess more advantages in terms of requiring less assumptions and being applicable to various situations [24]. As such, the primary problems involved in ENF based audio forensics are ENF detection and estimation. Therefore, ENF is a promising means for tackling audio forensic problems.

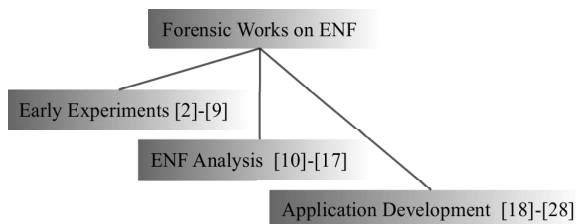


FIGURE 1. General categorization of existing works on ENF.

B. BRIEF REVIEW

The existing works on ENF based audio authentication system can be classified into three categories, i) early experiments and engineering [2]–[9], ii) ENF estimation and analysis [10]–[17], and iii) application development [18]–[28]. This is depicted in Fig. 1. One of the earliest works that discover the possibility of using ENF as an audio forensic criterion dated back to 2005 [2], followed by its expanded work in [3], where the experiments were conducted in Romania. Following this pioneering work, experiments were then carried out in several other European countries, including Poland [4], the United Kingdom [5], Netherlands [6], and Macedonia [7], and then in north America [8], not only in AC powered devices, but also in battery powered ones [9]. Among the experimental works, the core system structure had been well created in [3] and then was commonly incorporated in later works till today. In general, the system consists of two major processing flows, dedicated to testing data and reference data processing respectively, where the outputs meet at a matching module to generate the timestamp information. In [5], an excellent description of the system with several additional minor components, e.g., interpolation and median filtering, was provided. Therefore, a typical ENF based audio timestamp verification system that embodies the aforementioned core structure can be designed as shown in Fig. 2 [16], where the median filtering process is optional. The matching process is usually performed exhaustively, using minimum mean squared error (MMSE) or maximum correlation coefficient as the condition to select matched index. We believe that the system in Fig. 2 is to date still the most recommended system to perform audio evidence timestamp verification for its very satisfactory performance and being easy to implement. Note that the ENF signal recorded in the reference database is

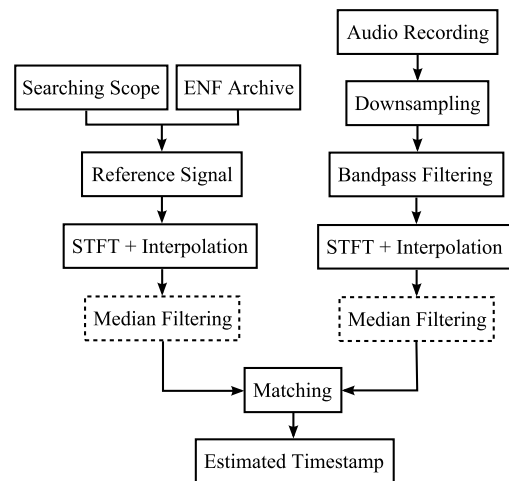


FIGURE 2. ENF based audio timestamp verification system structure [16].

usually assumed to contain the ground truth information, but in fact the reference signal is prone to errors if it is obtained from a single reference location [32]. This is not in the scope of this paper, and the readers are directed to [32] on how to build more reliable ENF database exploiting geographical information and harmony search.

Since the extraction of ENF signal from audio recordings is a classical spectral estimation problem, much research efforts have been spent on more accurate ENF estimations. Statistical modeling is applied to ENF analysis in [10]. A nonparametric adaptive source localization technique and a frequency tracking method based on ENF's slow-varying nature are used in [11]. The fact that ENF signals exist in audio recordings in terms of harmonics has been exploited in [12] and [13]. In [14], an improved discrete Fourier transform (DFT) based method that searches through the spectral lines within a reasonable ENF band is proposed. High precision phase analysis is proposed in [15] as another alternative means to extract ENF signals. A dynamic matching algorithm is proposed in [16], where frequency resolution of short-time Fourier transform (STFT) is used to reasonably correct deviated instantaneous frequencies. Finally, parametric methods for ENF estimation are investigated in [17] for improved frequency estimation.

Despite the intensive focus on ENF estimation and the application of timestamp verification (as in the studies summarized above), researchers have been looking into possible alternatives to diversify the applications of the ENF. One meaningful attempt is to use ENF to geolocate audio recordings [18], based on a machine learning framework that explores the differences among the ENF signals in different countries. With the same objective, in [19], the authors have shared a novel perspective of regarding ENF extraction as a frequency demodulation problem, where off-the-shelf methods can be applied to extract location-unique features. ENF security and countermeasure issues are investigated in [25], where the detection of deliberate ENF removal and insertion are investigated. Another interesting and well developed work

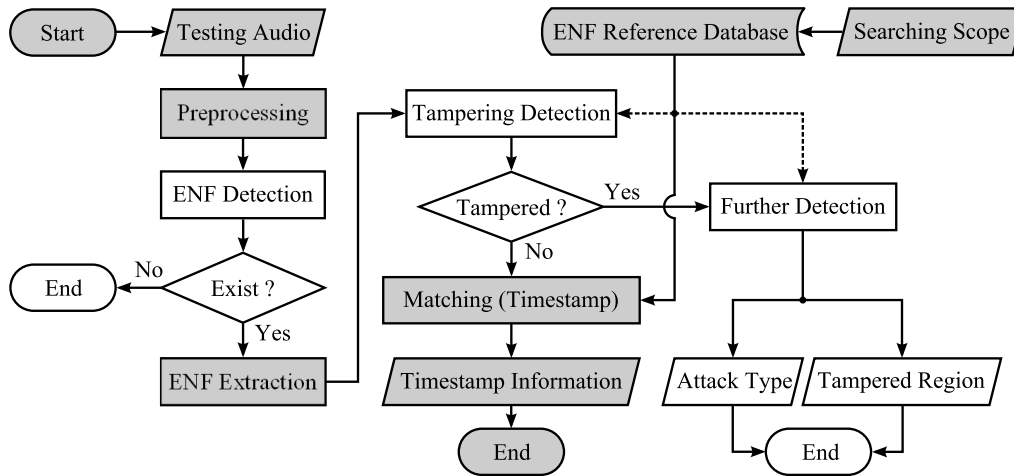


FIGURE 3. Flowchart of a practical ENF based audio authentication system.

explores the existence of ENF in indoor lighting systems, which opens the door of the investigation of ENF signals in videos [26]. A new application that utilizes the ENF signal to synchronize multimedia content is proposed in [27]. In recaptured audio recordings, there could exist two versions of ENF signals, one from the originally captured audio and the other during the process of recapturing. ENF analysis and decoupling in such recordings are investigated in [28]. Besides the above, some research attentions have been drawn on whether ENF could be used for audio tampering detection [15], [20]–[24]. However, we are still in an early stage in the development towards a systematic solution that requires neither human interaction nor less-practical assumptions.

While, as briefly summarized above, many works have been done to improve the accuracy of ENF estimations and discover novel ENF based applications, it is important to think about a primary question: how reliable is the ENF criterion in audio forensic investigations? In this paper, we look into several practical issues of ENF based audio forensic analysis systems. Specifically, we comprehensively address the problems that ENF analysts and system designers are currently facing. These problems are i) on the existence of ENF in audio recordings, ii) limitations of ENF based audio tampering detection systems, and iii) noise and interference issues, respectively. Because the research on the ENF criterion is highly application driven (originally, audio evidence timestamp verification), clarifying what the practical situations of the application are and what solutions are needed is an important issue. Unfortunately, this has not been adequately addressed in current literature. On the other hand, we also aim to shed some light on how to more effectively make use of the ENF as an audio forensic criterion to facilitate more reliable applications.

The paper is organized as follows. Section II presents a big picture of practical ENF based audio authentication systems, and then discusses ENF detection problems. Section III addresses ENF based audio tampering detection problems,

reviewing existing solutions and commenting on the limitations. In section IV, we illustrate the difficulties in ENF analysis, and discuss the noise and interference that affect ENF estimation performance. Strategies towards reliable ENF based audio authentication systems are investigated in Section V. Conclusion is made in Section VI.

II. ON THE EXISTENCE OF THE ENF

A. THE BIG PICTURE

As an application driven research topic, it is essential to understand the application specifications and requirements before conducting targeted research. Existing literature is usually based on one or several, explicit or implicit assumptions about the ENF and conditions of audio recordings, e.g., the ENF signal is surely captured, noise and interference do not pollute the frequency band of ENF dynamics, the audio recordings have not been tampered with, etc. As a result, the applicability of a solution may be limited. Further, it becomes more and more ambiguous of what the application would generally look like in practice. Before we start detailed discussions, the big picture is rigorously provided in this subsection. Here, we focus on two mostly investigated applications, i.e., timestamp verification and tampering detection, which is simply known as audio authentication.

The practical situation for an ENF based audio authentication system is demonstrated in Fig. 3. In early development stages [2]–[9], the major contribution is in terms of revealing and establishing the minimally required components which are marked grey in Fig. 3. The processing flow associated with the grey components is also a general picture of Fig. 2. A large amount of the further research and development works [10]–[16], [19] are on the grey component of “ENF Extraction”, a classical problem of single tone spectral estimation in noisy observations [33], [34], rather than developing other components for practical robustness. As a result, the problems of ENF detection and tampering detection have not been studied as sufficiently as ENF extraction. If any

uncolored component encounters a problem, then the “grey” system is going to produce a wrong output. Imagine that a testing audio recording contains corrupted (in whatever ways) ENF information, then following the grey processing flow will still yield an estimated timestamp but with large matching errors. No insights could be drawn from such a result in determining whether the match is trustworthy or the searching scope contains the true timestamp of the testing audio recording. For the issue of tampering detection, problems become more complicated because of the unknown tampering attack type and tampered region. Basically, one may need to first perform tampering detection, and then detect the attack type and tampered region if necessary. Currently, ENF based audio tampering detection still remains an open problem. Note that the reference data can be optional to facilitate tampering detection (e.g., used in [21], not used in [15]), which is shown as dashed lines in Fig. 3. In the remainder of this section and the sections followed, we will intensively discuss the problems of ENF based audio authentication systems. Each of the problems is an obstacle in front of one or multiple components in Fig. 3.

B. ENF DETECTION PROBLEMS

Let us begin with the very initial stage of the big picture, i.e., an audio recording has been made. Several problems arise at this stage. First, without any doubt, the purpose of creating an audio recording is to record some content of information (speech, music, etc.) and ENF is usually captured unintentionally. Thus, the distance between the recording device and nearby power sources does not have a guarantee. Note that the strength of an acoustic signal is inversely proportional to the square of the distance between the signal and the sensor. Meanwhile, the strength of ENF hum is always much weaker than the recorded content of interest, even the recording is made in good proximity of power outlets. We can therefore infer that on average, the captured ENF signal is indeed very weak. This is especially problematic when the audio content occupies the frequency band of ENF signals. In-band noise and interference will be discussed with more details in Section IV-B. Researchers and practitioners are aware of these problems, as reflected from the existing literature. However, there exists another inevitable yet unreported problem about ENF detection, that is, even in audio recordings made in environments without ENF, some signal can still be extracted from the “ENF Extraction” component in Fig. 3.

To illustrate ENF detection problem more systematically, we denote the testing audio samples in time domain as \mathbf{y} , which could be the superposition of several components. If all components are captured, then we have

$$\mathbf{y} = \mathbf{s} + \mathbf{x} + \mathbf{n}, \quad \mathbf{y}, \mathbf{s}, \mathbf{x}, \mathbf{n} \in \mathbb{R}^{N \times 1}, \quad (1)$$

where \mathbf{s} is the ENF signal, \mathbf{x} is the audio content, \mathbf{n} is the background noise, and N is the length of the signal. While \mathbf{x} surely exists in this model, the existence of \mathbf{s} and \mathbf{n} may not be guaranteed. For example, in a recording of an interview conducted in a quiet office room, the background noise \mathbf{n} can

TABLE 1. Possible Signals Captured in Audio Recordings

	ENF Captured	ENF not Captured
Noise Free	$\mathbf{s} + \mathbf{x}$	\mathbf{x}
Noisy	$\mathbf{s} + \mathbf{x} + \mathbf{n}$	$\mathbf{x} + \mathbf{n}$

be ignored, while in a recording made outdoor, \mathbf{s} may not exist in surrounding area. All the possible situations on the existence of ENF in audio recordings are summarized in Table 1. In a typical ENF analysis system with the assumption that \mathbf{s} is captured, the task is to extract \mathbf{s} from the mixture $\mathbf{s} + \mathbf{x}$ or the noisy version $\mathbf{s} + \mathbf{x} + \mathbf{n}$. Problems may arise when either \mathbf{x} or \mathbf{n} spans the frequency band of \mathbf{s} , but such problems are traceable and have been well studied. For example, the Cramer-Rao lower bound (CRLB) [33], [34], can well quantify the errors for different noise levels.

However, if \mathbf{s} has not been captured in an audio recording, and its frequency band is contaminated by \mathbf{x} or $\mathbf{x} + \mathbf{n}$, then the output of the right hand side processing branch in Fig. 2 or the grey processing components from “Start” to “ENF Extraction” in Fig. 3 will be nontrivial and contain some signal. Without any prior knowledge of an audio recording, the analyst is not able to determine whether Fig. 4 (b) corresponds to only noise and audio content or ENF signal with strong noise. Let the downsampled and bandpass filtered signals be $\tilde{\mathbf{y}}$, $\tilde{\mathbf{s}}$, $\tilde{\mathbf{x}}$, and $\tilde{\mathbf{n}}$ in the vector space of $\mathbb{R}^{M \times 1}$, where M is associated with N by the resampling ratio. A rigorous formulation of ENF detection problem gives the following hypotheses

$$\begin{aligned} G_0 & \begin{cases} H_0 : \tilde{\mathbf{y}} = \tilde{\mathbf{x}}, \\ H_1 : \tilde{\mathbf{y}} = \tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \end{cases} \\ G_1 & \begin{cases} H_2 : \tilde{\mathbf{y}} = \tilde{\mathbf{s}} + \tilde{\mathbf{x}}, \\ H_3 : \tilde{\mathbf{y}} = \tilde{\mathbf{s}} + \tilde{\mathbf{x}} + \tilde{\mathbf{n}}. \end{cases} \end{aligned}$$

Note that there is no loss of information when converting \mathbf{s} to $\tilde{\mathbf{s}}$. The above problem is generally hard because $\tilde{\mathbf{s}}$, $\tilde{\mathbf{x}}$, and $\tilde{\mathbf{n}}$ occupy the same frequency band, and the statistics of \mathbf{x} is unknown. One may simplify the problem to a binary decision problem by grouping the hypotheses into G_0 and G_1 . As a forensic problem, the solution of the detection problem should minimize the probability false positive decisions, i.e., $P\{G_1; G_0\}$, but little efforts have been made towards this objective. A possible approach to ENF detection could be observed from Fig. 4. It can be seen that when ENF is the dominant signal in the frequency band of interest, a continuous and slowly varying pattern exists in the time-frequency domain, while the signal contributed by \mathbf{x} and \mathbf{n} represents a discontinuous and random pattern. However, Fig. 4 (b) only shows a single example of audio recording without ENF. To rigorously develop the solution, large amount of recordings with and without ENF need to be created, and the corresponding time-frequency patterns within the frequency band of interest need to be studied. Feature extraction and

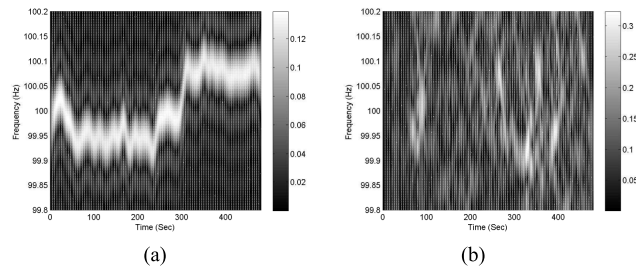


FIGURE 4. Illustrative examples of the results obtained from downsampling (to 400 Hz), bandpass filtering (passing 100 Hz), STFT (16 seconds window, 80000 sample FFT), and interpolation [35] applied to recordings (a) with and (b) without ENF signal captured. In (a), x does not span the frequency band of s . In (b), x and n span the frequency band of s .

machine learning techniques may be applied here. Alternatively, statistical analysis and hypothesis testing techniques could also be applied for effective solutions. We consider this problem as a good research direction.

III. ON ENF BASED TAMPERING DETECTION

A. IMPORTANCE OF TAMPERING DETECTION

In digital forensics, the integrity of an evidence is essential to determine its admissibility to investigation and law enforcement agencies. Thus, we believe that ENF based audio tampering detection takes a higher priority as compared to timestamp verification. There are two major reasons for this argument. i) If an evidence audio recording has been tampered with, then the embedded ENF signal would also be altered, which is very likely to lead to a wrong matched location to the reference data. Fig. 5 shows a typical example. The ENF of a testing audio recording is extracted and matched to the correct location of the reference data, i.e., index of 1501. However, if a portion of the testing data is deleted, then the matched location changes to 131, a very wrong match. Note that there exist several tampering attacks and each can be applied for multiple times on an audio recording. Therefore, if the testing file has been tampered with, in most situations the matching process will yield a wrong index. ii) If in some way the timestamp of the tampered data happens to be correctly verified, then a tampered file could become an evidence taking legal effects, which can cause serious consequences. Therefore, a tampering detection module is preferred to be deployed before timestamp verification, as shown in Fig. 3.

B. EXISTING SOLUTIONS AND LIMITATIONS

The detection of tampered multimedia content is generally a very difficult problem for any multimedia format. Among the existing works on ENF, audio tampering detection has been discussed in [15], [20]–[24]. We can group these works into two classes. The first class represents an intuitive idea of using ENF for tampering detection [20], [21], that is, with the assumptions of the true timestamp being known by some means and the ENF being dominant in its frequency band, the extracted ENF signals from the testing and reference data can then be placed

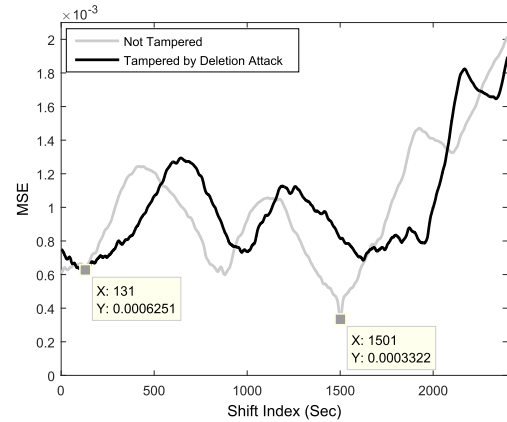


FIGURE 5. MSE curves during ENF matching with and without tampering attack.

side by side for a direct comparison, and the discrepancies can be explored to identify tampered regions. In practice, such an approach is feasible if the timestamp can be authenticated by other means (e.g., having witness, monitored by authority, etc.).

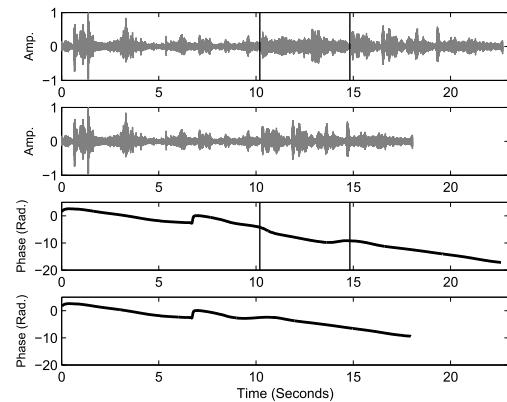


FIGURE 6. ENF discontinuity is unreliable for tampering detection [24].

However, more commonly, the true timestamp is unknown. In this situation, tampering detection solutions usually rely solely on the feature of the extracted testing ENF signal from the tampered audio recording, without the use of the reference database [15], [21]–[23]. Such a “feature” can be unified as the following statement: tampering attacks will cause discontinuity in the time-frequency representation of ENF signals, which can be explored for tampering detection. Such discontinuity is interpreted as phase discontinuity in [15], [22], and amplitude discontinuity in [21], [23]. However, in our opinion, the discontinuity should not be accepted as a fact caused by tampering attacks. Instead, such a “feature” is rather another assumption, and the corresponding solutions are prone to high false positive rate. An example supporting this argument is provided in Fig. 6 [24]. The first subfigure shows the original audio waveform, and the portion between the bars is deleted, resulting in the waveform shown in the second subfigure. The phase curves of the ENF signal

extracted from the original and the tampered audio data are shown in the third and fourth subfigures. It is clearly shown that phase discontinuity occurs at around the 7th second (region not tampered with), not occurring at the true tampered region, 10th second. The same problem also exists for the works based on amplitude discontinuity [21], [23], because amplitude discontinuity also commonly exists in the ENF signals of audio recordings that have not been tampered with (see Fig. 9 in [5]). In [24], the authors presented a systematic approach to jointly verify the timestamp and perform audio tampering detection based on the proposed absolute-error-map (AEM). In this work, neither the true timestamp nor the assumption of discontinuity at tampered region is assumed available, while the reference data with a searching scope are known. Then, signal segments that have not been tampered with are fully utilized to produce multiple local matches with the reference data represented in the AEM, based on which tampering detection mechanisms are designed. It should be noted that the system in [24] can deal with insertion, deletion, and splicing attacks, but only for a single attack applied in a single location for an audio recording.

In fact, the authors of the works on ENF based audio tampering detection are aware of the limitations as can be seen from their discussions in [21], [23], [24]. Apart from the limitations of the existing solutions, challenges also come from the variety of tampering attacks. Specifically, we have to deal with not only multiple attack types, e.g., deletion, insertion, splicing, replacement, shuffling, etc., but also combinations of attacks. Sometimes, those attacks can also appear at multiple locations. Therefore, whether ENF can be used for effective audio tampering detection still remains uncertain, and the system proposed in [24] actually offers a good starting point rather than a truly reliable solution.

To improve the reliability and performance of audio tampering detection, one may resort to non-ENF based solutions, which could be based on exploring compression offsets [29], acoustic signature [30], or features applied to machine learning techniques [36], [37], etc. Alternatively, if ENF is still used, then the assumption on the discontinuity at tampered location should not be used as the only criterion. Following the rationale in [24], one may look into possible ways to fully utilize the available practical information provided by the testing audio recording and the reference data.

IV. ON ENF ANALYSIS FOR TIMESTAMP VERIFICATION

While, as discussed in previous sections, ENF detection and ENF based audio tampering detection have attracted less research attention, researchers and practitioners have put more efforts in ENF extraction from testing audio recordings [10]–[16], [19]. However, despite the variety of solutions to analyze the ENF signals, the improvements provided by the existing solutions are marginal as compared to the classical system proposed in [5]. The reasons are twofold. First, in high signal-to-noise ratio (SNR) situations, improvements of ENF estimation could only be marginal, no matter how advanced the technique is, and the improvements will have

trivial effects on ENF matching performance. This concern is detailed in Section IV-A. Second, in low SNR situations, the bound on estimation errors (e.g., CRLB) will be inevitably increased, and no matter how advanced the ENF estimation technique is, the errors cannot be avoided. This will be detailed in Section IV-B. In general, the existing works on ENF estimation are concerned with how the estimation errors versus SNR can be close to the CRLB, but an audio evidence timestamp verification system is concerned with whether and how the extracted ENF signal can be effectively matched to the reference data.

A. HIGH SNR ENF ANALYSIS

The extraction of ENF signal is a time-frequency analysis problem, where we face the classical trade-off between time and frequency resolutions. However, since audio evidence timestamp verification is more concerned with whether the extracted ENF can have a good match with the reference data, such a trade-off becomes less important as long as the reference and testing signals are transformed using the same settings. Using STFT for example, as long as the frame size, stepsize, and fast Fourier transform (FFT) length are the same for estimating reference and testing signals, then the conditions for a good match are satisfied. This is well evidenced by Fig. 7, where different frame sizes (namely, different accuracies) are used with other parameters identically set for a testing audio recording and its corresponding one hour reference. Median filter proposed in [10] is not used here so that the estimation accuracy can be more clearly studied. The reference signal is obtained by directly connecting the recording system to a power outlet where the SNR approaches infinity. Therefore, the reference ENF signals obtained with different frame sizes are very similar to each other, as shown in Fig. 7 (a). However, we can still observe the higher perturbations from the smoothness of the curves when a shorter frame size is used. For the testing audio, since the sound (including ENF, audio content, and background noise) is captured via acoustic paths, the estimation process becomes more sensitive to the frame size. In Fig. 7 (b), we can see that a frame size of 2 seconds is definitely not an appropriate setting for time-frequency analysis because it highly compromises frequency resolution. However, when it is matched with the reference data with the same frame size, the matched index is only 1 second earlier than the ground truth timestamp, which is shown in Fig. 7 (c). Meanwhile, all the other settings can identify the true timestamp index.

It is important to note that although all frame size settings in Fig. 7 can produce the correct matching result, a large frame size is preferred in case of raised noise and interference. Ideally, the best setting is such that the smoothness of testing ENF can be similar to the smoothness of the reference ENF. Using such a criterion, it can be seen that 8 and 16 seconds are better choices. The above analysis indicates that the improvement of ENF estimation accuracy in existing works is not going to alter the matching result. In particular, we can use the four sets of the ENF reference and testing pairs

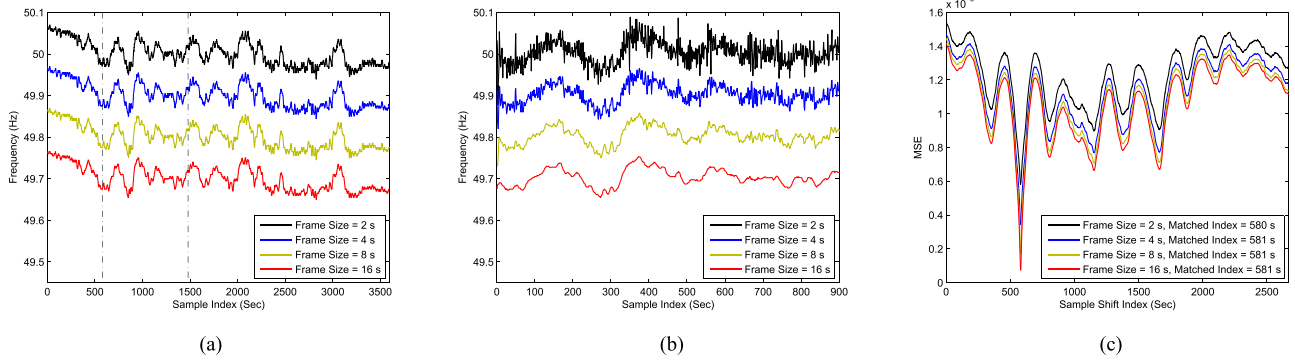


FIGURE 7. Demonstration of the trivial effects of ENF estimation accuracy on the matching output, where the reference (1 hour) and testing (15 minutes) signals are transformed using STFT and interpolation with different frame sizes. STFT stepsize is 1 second. In (a) and (b), the curves for frame sizes of 4, 6, and 8 seconds are shifted downwards by 0.1, 0.2, and 0.3 Hz respectively for comparison purpose. Curves in (b) are matched to the marked interval in (a).

in Fig. 7 (a) and (b) and compare them with Figs. 3 and 4 in [11], Fig. 6 in [12], and Figs. 11 and 13 in [14]. It can be seen that the improvement of estimation accuracy in these references is generally less than the improvement of changing frame size from 8 to 16 seconds in Fig. 7 of this paper, merely a marginal (trivial) improvement. Similar situations also occur in the phase analysis work [15], in which the improvements of phase analysis from Fig. 7 (c) to (d) or Fig. 8 (c) to (d) in [15] are very small, and we believe such improvements can be ignored without affecting the matching performance.

Another important issue in ENF based evidence audio timestamp verification is the duration of the testing recording. Usually, as can be seen from the majority of the existing works, the duration should be longer than 10 minutes. In [6], the authors defined a short recording as one with a duration less than 10 minutes. Furthermore, experiments on recordings of 2 minutes have been presented in [16]. Generally, the performance of ENF analysis degrades drastically when the duration of testing recording become less than 10 minutes, because the randomness (uniqueness) of the ENF sequence is weaker with decreasing lengths. Therefore, when dealing with short recordings, high accuracy ENF analysis becomes important. In view of this, future research attention should be put on improved ENF estimation with shorter frame sizes.

B. LOW SNR ENF ANALYSIS

The previous subsection has demonstrated that under good SNR conditions, the marginal improvement of ENF estimation accuracy is a trivial contribution to the matching output in general cases when dealing with long recordings. In this subsection, we will illustrate that when the noise and interference become stronger, improved ENF estimation accuracy could neither substantially reduce estimation errors nor essentially improve the matching result.

1) BACKGROUND NOISE

It is well known that the CRLB is the theoretical bound [33], [38] of the MSE that a minimum variance

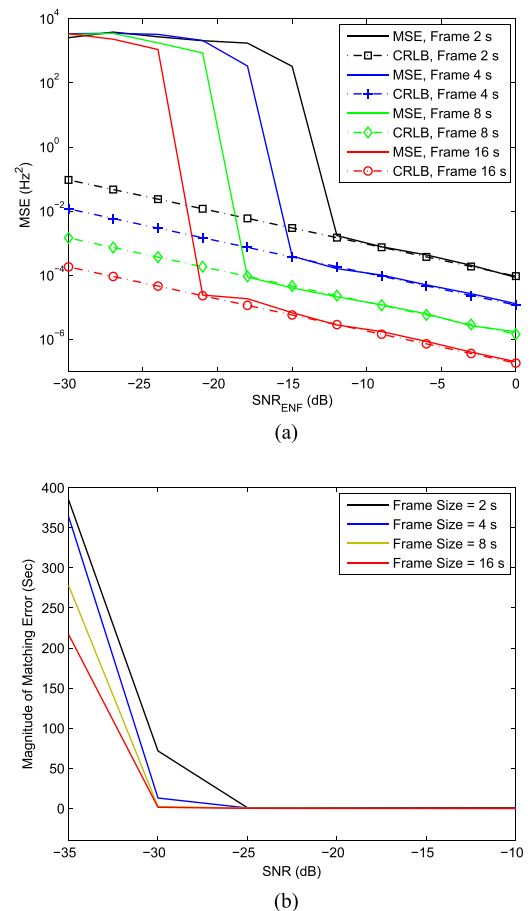


FIGURE 8. (a) Threshold effect of single tone estimation using DFT and (b) its impact on the ENF matching output, under different estimation accuracy settings. Curves are obtained by averaging 100 independent noise realizations. The matching performance exhibits a similar threshold effect, while the performance differences among different accuracy settings are trivial.

unbiased estimator (MVUE) can achieve. For ENF analysis, due to the fact that the captured ENF signal is usually much weaker than noise and audio interference, in most cases the analyst is dealing with moderate to low SNR situations.

In such situations, it is highly important to understand another fact that the threshold effect of a nonlinear estimator [33], [38] will result in inevitable estimation errors which could not be corrected no matter how advanced the estimator is. Here, we consider the ENF signal as deterministic signal corrupted by random noise as seen in most existing works, and CRLB applies to such a setting. To concertize the problem, we consider a synthetic model of (1) where \mathbf{n} is the synthetic independent and identically distributed (i.i.d.) white Gaussian noise (WGN). Then the original SNR is given by

$$\text{SNR} = 10\log_{10} \frac{\|\mathbf{s}\|_2^2}{\|\mathbf{n} + \mathbf{x}\|_2^2},$$

where $\|\cdot\|_2$ denotes ℓ_2 -norm. For simplicity and theoretical traceability, we do not consider the effect of \mathbf{x} here (will be considered in Section IV-B-2) of this subsection) and further define

$$\text{SNR}_{\text{ENF}} = 10\log_{10} \frac{\|\mathbf{s}\|_2^2}{\|\mathbf{n}\|_2^2},$$

which will be used for theoretical study of the performance of DFT for ENF analysis. According to [38], the CRLB of ENF estimation variance (ENF assumed to be deterministic but unknown) is given by

$$\text{CRLB}_{\text{ENF}} = \frac{12}{\text{SNR}_{\text{ENF}} \times N(N^2 - 1)} \times \left(\frac{f_s}{2\pi}\right)^2 \text{ Hz}^2,$$

where f_s is the sampling frequency. The threshold effects due to background noise are illustrated in Fig. 8. According to this figure, we can roughly quantify the high SNR situations discussed in Section IV-A as -10 dB and above, while the low SNR region could be -20 dB and below. Several important observations are noted in Fig. 8 (a). i) Increasing the frame size can alleviate the threshold effect so that stronger noise can be tolerated. Thus, it is recommended to sacrifice some time resolution during the calculation of STFT by using larger frame sizes. This is only feasible when the duration of testing audio recording is long enough. ii) In high SNR regions, we observe that the MSE curves are very close to the CRLBs, which indicates that DFT (used in each frame during STFT) is already a very promising means for ENF estimation. The argument in Section IV-A is hence supported by the observation here. iii) In the threshold effect region, the ENF could not be correctly estimated. In this situation, the existing works on improved ENF estimation also become ineffective.

Furthermore, the matching errors between the extracted ENF and the reference also exhibit a threshold effect, because the estimated ENF instantaneous frequencies are the input of the matching system. This is clearly shown in Fig. 8 (b). Note that Fig. 8 (a) is generated by synthetic signal without considering \mathbf{x} , while Fig. 8 (b) is based on a “silent” recording made in good proximity to ENF sources, which is then processed by adding different levels of WGN. Therefore, the role of \mathbf{x} can still be ignored. Also note that the theoretical study in Fig. 8 (a) did not consider downsampling and bandpass

filtering processes, but these processes have been applied in the example in Fig. 8 (b) to obtain final matching errors. Due to downsampling and bandpass filtering, the system can practically tolerate more noises. The fact that improved estimation accuracy is trivial in high SNR region and ineffective in threshold region is clearly shown in Fig. 8 (b). Unfortunately, based on our experiments, the noise conditions of practical recordings usually lie in moderate to low SNR regions, unless the recording device is placed next to a power outlet. Next, we discuss the interference term \mathbf{x} .

2) POTENTIAL INTERFERENCE

The above analysis based on the CRLB shows the unavailability of improving frequency estimation in low SNR conditions below the threshold when \mathbf{x} is assumed to be out of band. If \mathbf{x} also occupies the ENF bandwidth, then further performance degradation cannot be avoided. Therefore, audio content that is likely to occupy the frequency band of interest should be paid special attention to.

According to the type of audio content, the interference could come from human voice, music content, and special noises. The critical frequency spectral regions for ENF analysis, including several important harmonics, are 50 (or 60) Hz, 100 (or 120) Hz, and 150 (or 180) Hz, respectively, which have been commonly explored by existing works. In human voice, the lowest fundamental frequency can reach 80 Hz. Although the dominant frequency band of human voice is from several hundreds Hz to 2 kHz, it is not a guarantee that pure speech recordings will not affect the ENF captured alongside. Situation becomes more challenging when music is captured in audio recordings. The frequency band between 30 and 500 Hz is commonly used for bass and rhythm effects, which conflicts with the ENF band. The most common musical instrument, i.e., piano, has a lowest key of 27.5 Hz. Other than that, guitar, double bass, cello, etc., can all produce sound of frequencies at least less than 100 Hz. More commonly, electric bass systems have become very popular in modern music industry, which is considered to be a must-have in many genres of music. If a recording captures such sounds, then the ENF signal during the time when these sounds are captured is very likely to be destroyed. Noticeably, the degradation caused by musical instruments or sound effects do not permanently destroy the ENF signal because it makes no sense to have permanent bass in a music recording. Another type of interference is a series of special noises. For example, sound produced by friction or percussion is likely to affect the ENF band. However, this type of interference is not easy to be summarized. Here, we provide a typical example to illustrate this interference, which is shown in Fig. 9. This is a practical case we have encountered during the experiments. It can be seen that the creaking sound produced by wiping a white board could cause big errors during ENF extraction, and the erroneous interval is generally consistent with the time domain signal as shown by the two subplots. The error in ENF signal appears earlier than the time domain signal because a frame size of 8 seconds is used for ENF extraction.

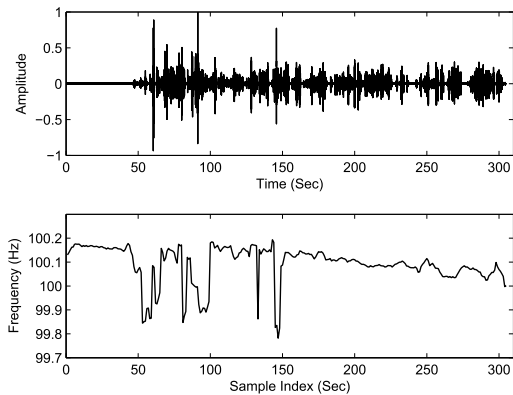


FIGURE 9. An example of special noise: creaking when wiping a white board. Upper: audio waveform. Lower: corresponding ENF signal extracted. The recording is a conversation between two persons in an office room. During 50 – 150 second, one speaker was writing on a white board and he wiped the white board for several times. The ENF in this interval was destroyed.

The conclusions about ENF extraction are summarized as follows. i) If the ENF is dominant in its frequency band, then the extraction process shown in Fig. 2 suffices to extract the ENF, while other methods may provide marginal improved accuracy but without changing the matching result, in most situations. ii) If the noise and interference are strong (especially beyond SNR threshold effect) in the frequency band of ENF signal, then large estimation errors are inevitable, and the system becomes unreliable because no method can break the CRLB. iii) The STFT based ENF extraction method is hence considered (also recommended) to be applicable to various situations according to i) and ii). iv) The use of a relatively large frame size for STFT is an effective solution for both ENF estimation and matching accuracies, which is reflected from both the experimental results and the $CRLB_{ENF}$ equation. v) More efficient techniques to deal with short recordings are worth of further investigations. vi) More research efforts should be devoted into the improvement of other components or the whole system in Fig. 3, rather than the ENF extraction component only.

C. SYNCHRONIZATION ISSUE

The synchronization issue in ENF based audio authentication systems refers to be the problem of slight clock desynchronization across many digital systems. For example, the clocks on the ENF reference database system, testing recording devices (iPhone, Samsung Galaxy, Sony Voice Recorder, etc.), and the IP Phone at our working location are all slightly desynchronized with inconsistency from several seconds to around 3 minutes. Even if a device is manually synchronized with a reference clock, it may also be out of synchronization due to hardware or software problems. During ENF extraction process, the processing granularity can be arbitrarily set according to practical needs. For example, all the ENF signals shown in this paper use a stepsize of 1 second, so that each sample corresponds to 1 second. Alternatively, we can modify the granularity less, or more than a second as

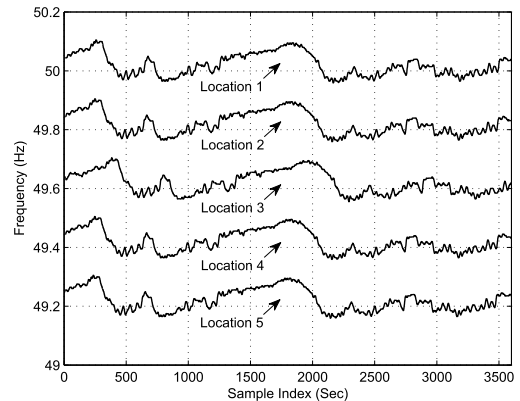


FIGURE 10. Clock desynchronization among five different ENF database locations. The curves for Locations 2 – 5 are shifted downwards by 0.2, 0.4, 0.6, and 0.8 Hz respectively for comparison purpose.

far as the sampling frequency permits. An example of the clock desynchronization among five different ENF database locations is provided in Fig. 10. One possible solution to this problem by synchronizing the database system with a GPS clock. Nonetheless, it really depends on how accurate the matching results are supposed to be, but if all the important clocks can be synchronized effectively, then the performance could be undoubtedly improved.

V. TOWARDS RELIABLE USE OF ENF

Some strategies to design more reliable ENF based audio authentication systems are briefly discussed in this section.

A. EXPLORING THE STRENGTH OF ENF

As shown in Fig. 3, ENF detection is an essential early step process in the audio authentication system. While it is rigorous to tackle this problem using signal processing and detection theory, which has been illustrated in Section II, it also should be noted that other means can be utilized to facilitate ENF detection from a forensic perspective. A primary question may arise from this perspective, that is, in what places can ENF be captured by recording devices? While the generic and conceptual answer that the recording device should be in the proximity of ENF sources has been known for a decade, it still lacks some practical guideline.

To work out a practical answer to this problem, extensive experiments are necessary, such as actively measuring the strength of ENF signals in various locations. The measuring device should have a sensor that only responds to ENF bandwidth. A slightly modified sound level meter can perform this task. In fact, the microphone on a smartphone can be effectively used as a sound level meter. The modification of the frequency response of the microphone can be carried out in software level. Specifically, one may design a few bandpass filters that pass ENF fundamental frequency and several harmonics, and then consolidate metering functions and compile the system as a smartphone application. If more accuracy is preferred, then one may resort to commercial products with customized specifications from sound

meter industry. We may call the device for measuring the strength of ENF as ENF level meter.

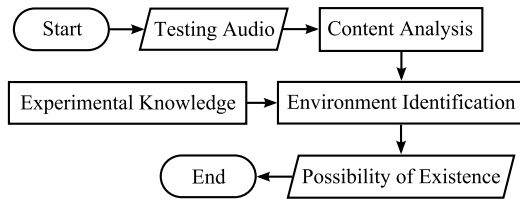


FIGURE 11. A qualitative pre-examination system on the existence of ENF.

With an ENF level meter, the strengths of ENF signals in various locations can be extensively measured. Potential locations include but are not limited to *cafes, restaurants, convenience stores, supermarkets, home rooms, quiet office rooms, office meeting rooms, classrooms, lecture halls, or even trains*. The selection of measuring locations is concerned with the size of the interior space, conditions of environmental noise, types of audio content that is likely to be recorded, and typical items placed in that place, etc. Sufficient experimental data can help us to gain the knowledge about how strong is the ENF signal in a specific or similar type of locations. Such experimental knowledge is especially helpful to law enforcement agencies when an audio evidence is known to be made in a location that has been measured. The qualitative pre-examination system that facilitates ENF detection is depicted in Fig. 11. In this figure, instead of being treated as interference from signal processing perspective, the audio content is actively analyzed to identify the location of the recording so that the experimental knowledge about ENF strength in that location can be used. Although it is an auxiliary support, important forensic clues can usually be obtained in practice.

B. STATISTICAL MODELING OF ENF

Currently, most existing works consider ENF as an unknown deterministic signal during the estimation process. Although it has been addressed in Section IV that ENF estimation accuracy is not vitally important, we may gain some insights by modeling ENF as an unknown random process and using it in ENF detection which has received less research attention. Therefore, statistical modeling of ENF may help to gain some *a priori* knowledge and may improve detection performance. Although not originally considered for detection purpose, statistical modeling has been noted in [10] where ENF is shown to follow a Gaussian distribution centered at the nominal frequency. Furthermore, an autoregressive (AR) model is proposed in [10] to model the ENF signal and hence improve the performance of timestamp verification. The slow-varying nature of ENF is well captured by the AR model therein. In addition, the authors in [11] have also utilized the slow-varying nature of ENF and proposed a frequency tracking system that can effectively avoid spikes and sudden changes in ENF estimates. If we apply the above prior knowledge to

ENF detection problem, then a more reliable detector could be obtained. For example, the *a priori* distribution of ENF can be used to design a maximum *a posteriori* probability (MAP) detector which is able to regulate its test statistics under noise and interference. On the other hand, the slow-varying property can be used to suppress false positive (ENF not captured but detected) rate when case like Fig. 4 is encountered, if a quantitative setting can be incorporated in the detection process. Meanwhile, it is also important to explore statistical features of audio content and background noise to facilitate the design of ENF detectors.

C. PRESERVING LOCAL MATCHES FOR TAMPERING DETECTION

Tampering detection is the most difficult problem in ENF based audio authentication systems due to the lack of practical clues of tampering attacks and the complication of the ways to perform attacks. Apart from the assumptions that can easily be violated in practical situations, another phenomenon is shared among the existing solutions, that is, they only focus on the locations that have been tampered with (discontinuity). In contrast, we believe that a truly reliable tampering detection system should not merely explore clues at tampered locations, but focus more on the locations that have not been tampered with, because the ENF segments in such locations can still be correctly matched to the reference data. The work in [24] has provided several positive preliminary results towards this direction. Generally, the challenge for ENF based tampering detection can be considered as the need for new matching methods, where the reference data should also be involved in such a designing process.

VI. CONCLUSION

Despite the use of ENF having been a significant development in audio forensics, practical issues have been less considered, and the big picture of ENF based audio authentication system is not given in the literature. This is seen from a series of assumptions made in existing works, which may not be available in practical situations. This paper takes a rigorous and comprehensive perspective and embodies all the related practical problems in a big picture of ENF based audio authentication system, i.e., Fig. 3. The problems are mainly concerned with three aspects. First, ENF detection has been shown to be a difficult problem because i) it is originally a multiple hypothesis testing problem involving three components \tilde{s} , \tilde{x} , and \tilde{n} , and ii) it is possible that a recording made in an environment without ENF can still have some signal lying in the frequency band of ENF dynamics (false positive decisions). Second, for ENF based audio tampering detection, current solutions are highly limited by the associated less practical assumptions such as true timestamp is known or tampering attacks will cause signal discontinuity. Problems become more difficult when multiple attacks, combination of attacks, and multiple attacked locations (which are quite often in practice) are conducted by attackers. Third, many research efforts have been devoted to the improvement

of ENF estimation accuracy, where ENF matching performance is marginally improved. We emphasize on the need for research and development aiming at improved ENF analysis with shorter frame sizes, which has not been well studied in the existing works. In general, majority of the existing works have focused on the “grey” system in Fig. 3 rather than its big picture. Future research efforts are recommended to be devoted to developing effective solutions to the other components in this big picture.

Some discussions on the means towards more reliable use of the ENF are also provided in this paper. To tackle the problems reported in Sections II-IV, possible approaches include but not limited to applying machine learning techniques to ENF detection problem, incorporating other tampering detection solutions and the reference database to facilitate ENF based tampering detection, and exploring more efficient timestamp matching modules. In terms of experimental and analytical works, it is worth of performing extensive measurements of ENF signal strengths in various locations, mining features of audio content and background noise, and modeling the ENF signal statistically.

REFERENCES

- [1] R. C. Maher, “Audio forensic examination: Authenticity, enhancement, and interpretation,” *IEEE Signal Process. Mag.*, vol. 2, no. 2, pp. 84–94, Mar. 2009.
- [2] C. Grigoras, “Digital audio recording analysis: The electric network frequency criterion,” *Int. J. Speech Lang., Law*, vol. 12, no. 1, pp. 63–76, 2005.
- [3] 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.
- [4] M. Kajstura, A. Trawinska, and J. Hebenstreit, “Application of the electrical network frequency (ENF) criterion: A case of a digital recording,” *Forensic Sci. Int.*, vol. 155, nos. 2–3, pp. 165–171, Dec. 2005.
- [5] 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, 2009.
- [6] M. Huijbregtse and Z. Geradts, “Using the ENF criterion for determining the time of recording of short digital audio recordings,” in *Proc. 3rd IWCF*, vol. 1, Aug. 2009, pp. 116–124.
- [7] 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. Elect. Electron. Equip.*, May 2012, pp. 1164–1169.
- [8] 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. General Meeting*, Jul. 2011, pp. 1–7.
- [9] J. Chai, F. Liu, Z. Yuan, R. W. Conners, and Y. Liu, “Source of ENF in battery-powered digital recordings,” *Audio Eng. Soc. Conv.*, vol. 135, pp. 1–6, Oct. 2013.
- [10] R. Garg, A. L. Varna, and M. Wu, “Modeling and analysis of electric network frequency signal for timestamp verification,” in *Proc. IEEE Int. WIFS*, Dec. 2012, pp. 67–72.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] D. P. N. Rodríguez, 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.
- [16] 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.
- [17] 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. (APSIPA ASC)*, Dec. 2012, pp. 1–10.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] J. Chai, L. Yuming, Z. Yuan, R. W. Conners, and Y. Liu, “Tampering detection of digital recordings using electric network frequency and phase angle,” *Audio Eng. Soc. Conv.*, vol. 135, pp. 1–8, Oct. 2013.
- [22] 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. (ChinaSIP)*, Jul. 2013, pp. 427–431.
- [23] P. A. Andrade 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] H. Su, A. Hajj-Ahmad, M. Wu, and D. W. Oard, “Exploring the use of ENF for multimedia synchronization,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Florence, Italy, May 2014, pp. 4613–4617.
- [28] 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. (ICASSP)*, May 2013, pp. 3018–3022.
- [29] R. Yang, Z. Qu, and J. Huang, “Exposing MP3 audio forgeries using frame offsets,” *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 8, no. 2S, pp. 1–20, Sep. 2012.
- [30] H. Malik, “Acoustic environment identification and its applications to audio forensics,” *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1827–1837, Nov. 2013.
- [31] L. Cuccovillo, S. Mann, M. Tagliasacchi, and P. Aichroth, “Audio tampering detection via microphone classification,” in *Proc. IEEE 15th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2013, pp. 177–182.
- [32] M. M. Elmesalawy and M. M. Eissa, “New forensic ENF reference database for media recording authentication based on harmony search technique using GIS and wide area frequency measurements,” *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 4, pp. 633–644, Apr. 2014.
- [33] D. Rife and R. Boorstyn, “Single tone parameter estimation from discrete-time observations,” *IEEE Trans. Inf. Theory*, vol. 20, no. 5, pp. 591–598, Sep. 1974.
- [34] S. M. Kay, *Fundamentals Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
- [35] E. Jacobsen and P. Kootsookos, “Fast, accurate frequency estimators,” *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 123–125, May 2007.
- [36] R. Korycki, “Time and spectral analysis methods with machine learning for the authentication of digital audio recordings,” *Forensic Sci. Int.*, vol. 230, nos. 1–3, pp. 117–126, Jul. 2013.

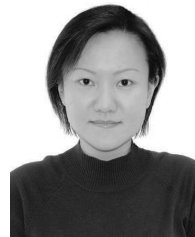
- [37] R. Korycki, "Authenticity examination of compressed audio recordings using detection of multiple compression and encoders' identification," *Forensic Sci. Int.*, vol. 238, pp. 33–46, May 2014.
- [38] S. M. Kay, *Fundamentals Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.



GUANG HUA 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 2013 to 2015, he was a Research Scientist with the Department of Cyber Security & Intelligence, Institute for Infocomm Research, Singapore. He then joined the School of Electrical and Electronic Engineering, Nanyang Technological University, as a Research Fellow. He is currently with the School of Electronic Information, Wuhan University. He has first authored over ten highly ranked IEEE journal papers. He holds a Singapore patent. He has a strong background in digital signal processing, especially for the processing of acoustic, radar, and multimedia signals, which cover a wide area of research interests, including array beamforming, digital filter design, applied convex optimization, data hiding, and multimedia forensics applications.



GUOAN BI received the B.Sc. degree in radio communications from the Dalian University of Technology, China, in 1982, and the M.Sc. degree in telecommunication systems and the Ph.D. degree in electronics systems from Essex University, U.K., in 1985 and 1988, respectively. Since 1991, he has been with the school of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His current research interests include DSP algorithms and hardware structures, and signal processing for various applications including sonar, radar, and communications.



VRIZLYNN L. L. THING received the Diploma degree in electrical, electronics, computer and communications engineering from Singapore Polytechnic, the B.Eng. degree (Hons.) and the M.Eng. degree in electrical, electronics, computer and communications engineering from Nanyang Technological University, Singapore, and the Ph.D. degree in computer science from the Imperial College London, U.K. During her career, she has taken on various roles with the key focus to lead and conduct cyber security research and development that bring a positive impact to the economy and society. She leads the Cyber Security & Intelligence Research and Development Department, Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. Her team of Scientists and Research Engineers focus on cyber security, digital forensics, and security analytics research and technology innovation. She is also an Adjunct Associate Professor with the School of Computing, National University of Singapore, and the School of Information Systems, Singapore Management University. He also holds the appointment of Honorary Assistant Superintendent of Police (Specialist V) with the Singapore Police Force, Ministry of Home Affairs. She also participates actively as the Lead Scientist of collaborative projects with industry partners and government agencies, and the Co-Director of the ST-InfoSec-A*STAR Cyber Security Joint Lab, the Sopra Steria-A*STAR Cyber Security Joint Lab, and the Custodio-A*STAR Cyber Security Joint Lab.

...