Check for updates

Received: 17 July 2020 | Revised: 25 February 2021 | Accepted: 15 March 2021

IET Signal Processing

DOI: 10.1049/sil2.12039

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Blackman–Tukey spectral estimation and electric network frequency matching from power mains and speech recordings

Georgios Karantaidis | **Constantine Kotropoulos**

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece

Correspondence

Georgios Karantaidis, Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece.
Email: gkarantai@csd.auth.gr

Funding information

General Secretariat for Research and Technology, Grant/Award Number: Unassigned, Greece; Hellenic Foundation for Research and Innovation, Grant/Award Number: Unassigned, Greece

Abstract

Forensic applications exploit electric network frequency (ENF) as a fingerprint to determine multimedia content authenticity, as well as the time and region of multimedia recording. ENF is present at a nominal frequency of 50/60 Hz and its harmonics. Strong interference due to speech content deteriorates ENF estimation accuracy. Herein, the authors propose a non-parametric approach for ENF estimation, which incorporates a customised lag window design into the Blackman–Tukey spectral estimation method. Leakage reduction is formulated as a problem of energy maximisation within the main lobe of the spectral window. The proposed approach is compared to state-of-the-art methods for ENF estimation. Maximum correlation coefficient and minimum standard deviation of errors are employed to measure ENF estimation accuracy. Hypothesis testing is performed to determine whether the improvements in ENF estimation accuracy of the proposed approach over the state-of-the-art methods are statistically significant. Experimental results and statistical tests indicate that the proposed approach improves ENF estimation against many state-of-the-art methods.

1 | INTRODUCTION

Multimedia content has begun to permeate many aspects of everyday life. From publicly available content to that which is strictly private, information diffusion has led to an overwhelming explosion in multimedia manipulation and alteration. Notwithstanding the indisputable multimedia forgery outbreak, many efforts have been paid to the development of efficient countermeasures detecting such criminal activities [1, 2].

Multimedia forensics acts crucially in recovering probative evidences from multimedia content and constitutes a powerful weapon in crime investigation. Multimedia authentication and integrity validation have become challenging tasks [3]. Moreover, the popularity of cloud computing has rendered it a target for criminals. To this end, great emphasis has been put on the development of efficient cloud forensics tools [4, 5].

Audio and image forensics are associated with a wide range of applications, for example, fraud detection, determination of region where recording took place, content authentication, and edit detection. Despite the significant progress in digital forensics, the proliferation of systems that edit audio recordings so that alterations are not noticeable constitutes a constant threat. Image forensics aims at providing tools for determining whether visual content has been modified or altered. Great emphasis has been put towards optimising their efficacy [6–9]. Copy-move attacks tend to be very common in the field of image forensics. Efforts have been made to develop robust techniques for their detection [10–12]. A robust method employing discrete cosine transform and singular value decomposition in order to detect this kind of attacks was presented in [13]. Robust scale and translation invariant features were employed to detect copy-move attacks. A block-based method employing Fourier–Mellin transformation for copy-move forgery detection was proposed in [14]. Watermarking has been used for image tamper detection. Many approaches based on watermarking have been proposed in order to determine whether an image is authentic or has been modified [15–18].

Audio forensics has been intensively studied [19–22]. Audio reverberation significantly affects the quality of an audio recording and small differences can indicate alterations in the original recordings. A forensic tool was presented in [23] that models reverberation. Background noise can be employed in order to determine the integrity of an audio recording, but

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Signal Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

speech leakage is evident. A novel framework was introduced for background noise estimation [24]. Splicing detection in audio recordings constitutes a major scenario in audio tampering detection. Convolutional neural networks (CNNs) are able to derive representative features of an audio recording. A novel method was proposed in [25] for splicing detection that employs CNN in order to derive such features using the spectrogram of audio recordings.

A groundbreaking tool in forensics is the electric network frequency (ENF) criterion proposed by Catalin Grigoras [26, 27]. By employing ENF, one can determine whether an audio recording has been modified or not, indicate when the recording occurred, as well as where the recording was captured. ENF is embedded in digital audio recordings and estimation techniques aim at extracting it in the optimal way. The ENF is present at both the nominal frequency and its harmonics. Having extracted ENF, it is compared against a reference database in order to determine its authenticity, the time the recordings were captured, and to indicate whether any alterations have occurred. Interferences and low signal-to-noise ratio (SNR) conditions hinder ENF estimation. Several approaches have been proposed to effectively extract ENF [28–34].

The design of a lag window for non-parametric Blackman–Tukey (BT) spectral estimator to optimally detect ENF in recordings from power mains as well as speech recordings is described herein. The designed window makes a compromise between smearing and leakage in order to detect ENF, whose power is weaker than that of other close frequency components of the signal (i.e. pitch or first formant of vowels). BT adopting the designed window offers more accurate detection in recordings from power mains and speech recordings compared to state-of-the-art methods due to the optimal separation between the ENF and the interferences. It is worth noting that the proposed approach, namely BT + LWD, is tested on real-world data sets in a compliant manner to the setup used in the literature. Two metrics are employed to assess ENF estimation accuracy, that is, the maximum correlation coefficient (MCC) and the minimum standard deviation of errors (mSDE) between the estimated ENF and the reference ENF. Furthermore, hypothesis testing is performed to assess whether the improvements in ENF estimation accuracy with respect to MCC and mSDE delivered by the proposed approach against the same metrics by state-of-the-art ones are statistically significant. It should be stressed that an estimation problem is addressed herein. That is, ENF is extracted using spectral estimation methods, when the reference ENF signal is available. The quality of the ENF estimation is assessed through matching, as is done in the related literature.

In summary, the main contributions described herein are as follows:

- A Blackman–Tukey approach is proposed, which employs a custom-designed lag window for efficiently estimating the power spectrum of the ENF component in recordings from power mains and speech recordings.
- A systematic study is conducted regarding parameter selection in designing the proposed lag window enabling accurate ENF estimation. Proper selection of parameters reduces leakage, affecting ENF estimation.
- Experiments are conducted on real-world benchmark data sets and extensive comparisons with state-of-the-art methods are made.
- Hypothesis testing asserts that the improvements in ENF estimation accuracy with respect to MCC and mSDE between the proposed approach and those disclosed in the literature are statistically significant.

1.1 | Related work

A non-parametric iterative adaptive approach (IAA) for spectral estimation was proposed in [28]. Furthermore, a novel method for frequency estimation based on dynamic programming, seeking the minimum cost path among per-frame frequencies was introduced also in [28]. Another approach was proposed employing the maximum-likelihood estimator (MLE) for multi-tone and single-tone signals [29]. In particular, the Cramer–Rao bound was used to estimate the variance of the ENF estimator. ENF is present at multiple harmonics of the nominal frequency. A method weighting ENF with the local SNR of each harmonic has been described [30]. An ENF estimation algorithm based on discrete Fourier transform (DFT) was proposed in [31]. There, a binary approach was used to seek specific spectral lines instead of the entire frequency band. Interference in speech signals hinders ENF estimation. To cope with that problem, robust principal component analysis was employed for noise reduction and weighted linear prediction for ENF estimation was introduced in [35]. Fine tuning of signal filtering and parametrisation for ENF estimation in recordings from power mains and speech recordings can increase ENF estimation accuracy. In this context, a systematic study of various techniques in combination with proper parametrisation was conducted in [36]. An ENF estimation method in audio recordings employing frequency demodulation was proposed in [37]. Different noise conditions were established in order to test different methods. A variety of parametric and non-parametric frequency estimation methods were employed for high-precision ENF estimation [38]. A method based on instantaneous frequency estimation using the Hilbert transform was proposed in [39]. Time requirements and estimation accuracy crucially affect ENF real-world applications. Window selection can significantly improve accuracy without affecting time requirements at all. A novel accurate approach employing a fast Capon-based spectral estimator after applying a temporal Parzen window was proposed in [40]. In most cases, the ENF signal in audio recordings suffers from strong interferences. Filtering may enhance the ENF signal in such cases. A filtering algorithm was proposed in [41], which employs a kernel function to create a time–frequency representation facilitating ENF estimation. The existence of reference ENF is of high importance in multimedia authentication tasks. A method for creating a

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

reliable ENF database, employing multiple frequency sensors, was detailed in [42]. An automated general scheme for ENF estimation was proposed in [34].

Machine learning algorithms can indicate the region where the recordings were captured [43]. A multi-label machine learning approach exploiting various ENF signal features to identify the region-of-recordings was examined [44]. CNNs can be exploited in order to learn features emerging from ENF audio recordings. A CNN-based system using spectrograms for audio recapture detection was proposed in [45].

ENF estimation accuracy in audio recordings is intuitively based on several stochastic and deterministic factors. The first comprehensive study on the nature of factors affecting ENF capture was conducted in [46]. Malicious attacks in power grids can be prevented when the future ENF is known. Two algorithms were proposed, employing correlation kernel regression and autoregressive moving average for ENF forecasting [47].

For completeness, tamper detection is discussed. ENF variations can be exploited to facilitate edit detection in multimedia recordings, as proposed in [48]. Phase discontinuities are created due to insertions and deletions. Phase change analysis via high-precision Fourier analysis was adopted to justify the authenticity of multimedia recordings [49]. A method employing an estimation of signal parameters with rotational invariant techniques (ESPRIT) and exploiting kurtosis for detecting abnormalities in ENF variations was suggested [50]. Time is critical in detecting multimedia alterations. A support vector machine (SVM)-based framework for automatic detection of such alterations based on the kurtosis of ENF disturbances was introduced [51]. Features of ENF signal were extracted and utilised for digital audio authentication. An SVM-based framework for automatic tamper detection with feature fusion was proposed in order to overcome difficulties in visual tamper inspection [52]. Finding edited areas constitutes a significant issue in digital forgery detection. A method based on the maximum offset in cross correlation between the estimated ENF and the reference one for edit detection was introduced in [53]. An algorithm based on inter-frame video forgery detection for frame deletion, duplication, and insertion has been proposed [54]. Timestamp verification and tamper detection were integrated within an audio verification system [55]. A measurement method was utilised based on absolute errors between the extracted ENF and the reference signal called Absolute-Error-Maps. ENF can be exploited as a fingerprint for audio and video synchronisation applications by aligning the embedded ENF signals. A scheme for multimedia synchronisation based on ENF was proposed in [56].

1.2 | Outline

Section 2 details ENF fundamentals, discusses ENF estimation, and describes the data sets employed in the experiments. Window design is analysed in Section 3. The experiments conducted to assess the proposed approach are discussed in Section 4, and the conclusions are drawn in Section 5. Notation and abbreviations are summarised in Table 1.

TABLE 1 Notation and abbreviations

r	Autocorrelation function
BT	Blackman–Tukey
$\hat{\phi}_{BT}$	Blackman–Tukey spectral estimator
$DTFT$	Discrete-time Fourier transform
ENF	Electric network frequency
$\mathbb{E}\{\}$	Expectation operator
C	Filter order
L	Frame length
FDR	Frequency disturbance recorder
Δf	Frequency resolution
$\mathbf{G}$	Ground truth data set
$\mathbf{W}$	Lag window
LWD	Lag window design
MCC	Maximum correlation coefficient
$mSDE$	Minimum standard deviation of errors
N	Number of data samples
K	Number of frames
PSD	Power spectral density
QI	Quadratic interpolation
F_s	Sampling frequency
$\mathbb{R}$	Set of real numbers
$STFT$	Short-time Fourier transform
SNR	Signal-to-noise ratio
S	Time shift
$\mathbf{A}$	Vector of DTFT

2 | ENF FUNDAMENTALS

2.1 | The ENF criterion

The instantaneous differences between the demanded and produced power generate stochastic fluctuations around the nominal frequency of ENF, which is at 50 Hz in Europe and 60 Hz in the United States [26]. These fluctuations are caused by the continuous alternations in rotational speed of energy generators in power plants. ENF can be recorded with specialised sensors called frequency disturbance recorders (FDRs), providing accurate measurements up to $\pm$ 5 $\times$ 10^{-4} Hz [28]. The most significant properties of the ENF signal are the following:

- ENF signal fluctuates in a random way around its nominal value.
- Identical ENF fluctuations are present within the same network.

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- Besides the nominal frequency, higher harmonics contain, also, the ENF signal [57].

The ENF criterion was introduced by Catalin Grigoras [26, 27] for determining the authenticity of digital recordings. Accurate methods should be developed in order to isolate and extract the embedded ENF traces and determine the integrity of multimedia content, the time of the recordings, as well as the region where the recordings were captured. Having extracted the ENF signal, a comparison against a reference ENF database should be made.

In Europe, ENF is defined as $f = [50 \pm \Delta f_E]$ Hz, where Δf_E denotes the ENF fluctuations around the nominal frequency [26]. Based on ENF fluctuations Δf_E, there are three different conditions, indicating the proper function of network operation [58]:

- If $\Delta f_E \leq 20$ mHz, the network is operating properly.
- If 20 mHz $< \Delta f_E \leq 200$ mHz, ENF fluctuations exceed normal limits, but there is no danger for network operation.
- If $\Delta f_E > 200$ mHz, a major risk in the electric network incurs.

Many factors should be taken into consideration to efficiently estimate ENF. Depending on the nature and properties of the ENF signal, three main categories of ENF estimation methods can be established based on the analysis domain [27]:

- Time/frequency: A visual comparison is made between the spectrogram and the reference signal. Usually, it is employed for short-time recordings.
- Frequency: The periodogram is computed and the frequency associated with its maximum magnitude (i.e. spectral peak) is estimated over short time segments. The corresponding frequency is compared against the reference ENF signal. This approach exploits various spectral estimation methods. A sharp bandpass zero-phase FIR filter should be applied on the raw signal, prior to spectral estimation.
- Time: Having applied a sharp bandpass zero-phase FIR filter to the raw signal, zero-crossing measurements around the frequency of interest (i.e. fundamental frequency or its harmonics) are conducted.

2.2 | Data set description and ENF estimation procedure

The presence of an ENF signal can be significantly affected by the environment, where the recordings are captured. Furthermore, an ENF estimation approach may be more efficient when tested on a specific recording environment. To this end, ENF estimation approaches must be tested on data sets based on various conditions and SNRs. Here, two benchmark data sets (www.sal.ufl.edu/download.html) provided by the University of Florida are employed in order to evaluate the proposed approach. These data sets consist of real-valued signals and were first discussed in [28] and have been employed also in [29, 35, 36, 40].

The first data set (Data 1) was recorded by connecting an electric outlet via a voltage divider directly to the internal sound card of a desktop computer. This is a high SNR data set, which indicates a strong signal. The second data set (Data 2) comprises speech recordings with low SNR, which indicates strong interference. This data set was captured by the internal microphone of a laptop computer. The original data sets were sampled at 44.1 kHz using 16 bits/sample. The initial recordings were downsampled at 441 Hz, using proper anti-aliasing filtering. Thus, apart from the fundamental frequency at 60 Hz, higher harmonics can be exploited for ENF estimation. Identical ENF signals should be contained in both data sets, since they were captured at the same time within the same interconnected network. Moreover, a reference ground truth data set containing the actual ENF signal was recorded by an FDR in order to be compared against the estimated signals from Data 1 and Data 2.

The first step in ENF estimation is bandpass filtering of the raw signal around the frequencies of interest. Since ENF is present at the fundamental frequency and its harmonics, the same procedure was performed in each case. During filtering, proper selection of both bandpass edges and filter order can significantly boost ENF estimation accuracy, as shown in [36]. Regarding Data 1, the first, second and third harmonics were employed in order for the experiments to be consistent with the literature. Around the first harmonic, a sharp zero-phase FIR filter with bandpass edges set at 59.9 and 60.1 Hz and filter order 1501, was applied. The bandpass edges for the second harmonic were set at 119.98 and 120.02 Hz, maintaining the same filter order as for the first harmonic. The third harmonic bandpass edges were set at 179.9 and 180.1 Hz. The filter order was decreased to 1001. In all cases, a Hamming window with length equal to filter order was employed. Regarding Data 2, the second harmonic was exploited only in the experiments, as done in the literature, since the other harmonics suffered from extremely low SNR. A sharp zero-phase FIR filter was applied around the second harmonic of the speech recordings with bandpass edges set at 119.9 and 120.1 Hz. The filter order was set at 4801. It is worth mentioning that selection of filter order affects critically ENF estimation accuracy. A different filter order needs to be defined at each harmonic to obtain the best results. As the filter order was increased, the corresponding transition width was reduced, yielding more accurate results. As a matter of fact, for the low SNR Data 2, the filter order was set at 4801, while for the high SNR Data 1, the filter order was set at 1001. Afterwards, the signal was split into V overlapping frames. Each frame of size L was shifted by S s from its immediate predecessor frame and was multiplied by an L-size temporal window. Here, the best results for each harmonic in Data 1 and Data 2 were derived employing an L-size Parzen window [59, 60] and an L-size rectangular window, respectively. Frame lengths for Data 1 and Data 2 are indicated in Table 2. The choice of temporal window is of crucial importance in ENF estimation. In [40], extensive experiments were conducted in order to determine the best window choice. Next, the power spectrum is estimated

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Frame parameters (s)

Parameters	Data 1	Data 2
Time shift, S	1	1
Frame length, L	20	33

for each frame and an approximate frequency ω_{qmax} associated with the maximum power spectrum magnitude is extracted. In order to derive a more precise frequency estimation, a quadratic interpolation is employed. Thus, a quadratic model is fit to the logarithm of the estimated power spectrum [28, 61].

3 | METHODS

3.1 | Blackman–Tukey

Let $\mathbf{y}_N = [y(t), \ldots, y(t+N-1)]^\top$ be the j-th data segment containing the raw samples after bandpass filtering and $(\cdot)^\top$ denote transposition. The autocorrelation between real-valued $y(t)$ and $y(t-n)$ is defined as: $r(n) = \mathbb{E}\{y(t)y(t-n)\}$, where $\mathbb{E}$ denotes the expectation operator.

The standard biased estimate of $r(n)$ is given by [62]:

$$\widehat{r}(n) = \frac{1}{N}\sum_{t=n+1}^{N} y(t)y(t-n), \qquad 0 \le n \le N-1 \tag{1}$$

For negative lags, $\widehat{r}(-n) = \widehat{r}(\mathrm{n})$. Classic periodogram-based methods suffer from high variability due to the accumulation of estimation errors. The Blackman–Tukey method yields a refined periodogram [62]:

$$\widehat{\phi}_{BT}(\omega) = \sum_{\zeta=-(M-1)}^{M-1} w(\zeta)\,\widehat{r}(\zeta)\,e^{-i\omega\zeta} \tag{2}$$

where $w(\zeta)$ is an even lag-window function. Here, M is equal to the number of samples N. For each data segment, $N = L\,F_s$, where F_s is the sampling frequency. For increased resolution, dense frequency samples are employed as $\omega_\xi = \frac{2\pi}{\Xi}\xi$, $\xi = 0, 1, \ldots, \Xi - 1$, where $\Xi = 4N = 4LF_s$.

Proper selection of the lag window may lead to an improvement of ENF estimation accuracy. Depending on the nature of the application, the best window choice will yield the desirable results. For each frame, the power spectral density (PSD) (2) attains a maximum at $\xi \in [0, \frac{\Xi}{2} - 1]$ or $f_\xi = \frac{\omega_{\xi\max}}{2\pi}F_s$. The computational complexity of the Blackman–Tukey spectral estimator is about $\Xi\left[\frac{1}{2}\log_2(\Xi) + 2\log_2(2\Xi)\right]$ [62], which is not prohibiting for large data sets.

3.2 | Lag window design

Taking into account the noisy nature of ENF signal, a lag window is designed, coined as LWD, in order to reduce interference that occurs in ENF signals and increases the ENF estimation accuracy. The lag window will be used in the BT spectral estimator. Leakage reduction is the main objective of lag window design. The trade-off between leakage and smearing should be taken into consideration for the selection of window shape. In ENF applications, especially when it comes to speech recordings, strong interference may mask the ENF signal through leakage. To this end, the design parameter θ employed herein, compromises smearing for reducing leakage as much as possible.

Let a discrete impulse response $\mathbf{d} = [d(0), \ldots, d(M-1)]^\top$ and

$$\mathbf{a}(\omega) = \left[1, e^{-i\omega}, \ldots, e^{-i(M-1)\omega}\right]^\top \tag{3}$$

be the vector of discrete-time Fourier transform (DTFT). DTFT of $\mathbf{d}$ can be written as

$$D(\omega) = \mathbf{d}^*\mathbf{a}(\omega) \tag{4}$$

where $(\cdot)^*$ stands for the Hermitian transposition of complex-valued vectors, following the notation in [62]. The spectral window is derived as [62]

$$W(\omega) = |D(\omega)|^2 \tag{5}$$

where $D(\omega)$ is the frequency response of any window.

The corresponding positive semi-definite lag window is given by the auto-correlation of $\mathbf{d}$, that is, for $-(M-1) \le \zeta \le (M-1)$,

$$w(\zeta) = \sum_{k=0}^{M-1} d(k)d^*(k-\zeta) \tag{6}$$

where * for scalars stands for conjugation.

Let $W(\omega)$ be the frequency response of the window function given by Equation (6). The following maximisation problem should be solved in order to maximise the relative energy in the main lobe of window $\mathbf{W}(\omega)$ [62]:

$$\max_{\mathbf{d}} \left\{ \frac{\int_{-\theta\pi}^{\theta\pi} W(\omega)\,d\omega}{\int_{-\pi}^{\pi} W(\omega)\,d\omega} \right\}. \tag{7}$$

For accurate ENF estimation, the choice of the window design parameter θ should provide the best trade-off between leakage and spectral resolution. By increasing θ, leakages will be mitigated at the expense of reduced spectral resolution. Here, $\theta = 3.5/M$ and $\theta = 2.1/M$ are employed for Data 1 and Data 2, respectively. Taking into account that

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\frac{\frac{1}{2\pi}\int_{-\theta\pi}^{\theta\pi}|D(\omega)|^2\,d\omega}{\frac{1}{2\pi}\int_{-\pi}^{\pi}|D(\omega)|^2\,d\omega}=\frac{\mathbf{d}^*\left[\frac{1}{2\pi}\int_{-\theta\pi}^{\theta\pi}\mathbf{a}(\omega)\mathbf{a}^*(\omega)\,d\omega\right]\mathbf{d}}{\mathbf{d}^*\mathbf{d}} \tag{8}$$

the maximisation problem (7) can be rewritten as the maximisation of Rayleigh quotient:

$$\max_{\mathbf{d}}\frac{\mathbf{d}^*\mathbf{\Psi}\mathbf{d}}{\mathbf{d}^*\mathbf{d}} \tag{9}$$

where $\mathbf{\Psi}$ is the Toeplitz matrix

$$\mathbf{\Psi}=\frac{1}{2\pi}\int_{-\theta\pi}^{\theta\pi}\mathbf{a}(\omega)\mathbf{a}^*(\omega)\,d\omega\overset{\Delta}{=}[\psi_{m-n}] \tag{10}$$

with elements

$$\psi_{m-n}=\frac{1}{2\pi}\int_{-\theta\pi}^{\theta\pi}\mathbf{e}^{-i(m-n)\omega}\,d\omega=\theta\mathrm{sinc}[(m-n)\theta\pi]. \tag{11}$$

In Equation (11), the sinc function is defined as $\mathrm{sinc}(x)\overset{\Delta}{=}\frac{\sin x}{x}$. Thus, $\mathbf{\Psi}$ is a real symmetric Toeplitz matrix. Accordingly, its eigenvectors have real entries.

The optimal lag window, which maximises the relative energy in the main lobe of $W(\omega)$ is obtained, when $\mathbf{d}$ is chosen as the principal eigenvector of $\mathbf{\Psi}$ [62]. Here, $\mathbf{d}$ is a real vector. The best results of each harmonic with respect to MCC and mSDE are derived employing $M=0.25LF_s$ and $M=0.5LF_s$ for Data 1 and Data 2, respectively. The window function $w(\zeta)$ and its corresponding frequency response are depicted in Figures 1 and 2, respectively.

3.3 | Quadratic interpolation

In order to calculate a more precise estimate of ENF, quadratic interpolation (QI) is employed. Thus, a quadratic model is fit to the logarithm of the estimated power spectrum about $\omega_{\xi\max}$ [34, 61]. QI enables high-resolution ENF estimation in combination with low time requirements. The frequency sample $\omega_{\xi\max}$, which corresponds to the maximum value of spectral magnitude is extracted as an approximate ENF estimate for each frame. The procedure is briefly described, as proposed in [34]. Let $\lambda_\alpha=\log\widehat{\phi}_{BT}(\omega_{\xi\max+\alpha})$, $\alpha=-1,\,0,\,1$, for each frame:

- Select the bin of the maximum power spectrum, $\lambda_0=\log\widehat{\phi}_{BT}(\omega_{\xi\max})$.
- Select the two adjacent bins of λ_0, that is, $\lambda_{-1}=\log\widehat{\phi}_{BT}(\omega_{\xi\max-1})$ *and* $\lambda_1=\log\widehat{\phi}_{BT}(\omega_{\xi\max+1})$.
- Calculate the quadratic peak δ, which corresponds to an improved ENF estimation.

The improved ENF estimate is obtained as $\omega=\omega_{\xi\max}+\delta$, where

$$\delta=\frac{1}{2}\frac{\lambda_{-1}-\lambda_1}{\lambda_{-1}-2\lambda_0+\lambda_1}(\omega_{\xi\max+1}-\omega_{\xi\max}) \tag{12}$$

3.4 | Accuracy of ENF estimation

After ENF estimation, a matching procedure is applied in order to objectively assess estimation accuracy. Having a ground truth (reference) data set, two metrics, namely, the MCC [33] and the mSDE [28, 63], are used to compare the extracted frequencies against the ground truth ones. Using the notation introduced in [28], let $\mathbf{f}=[f_1,f_2,\ldots,f_K]^\top$ be the estimated ENF signal at each second. Let also $g=[g_1,g_2,\ldots,g_{\tilde{K}}]^\top for \tilde{K}>K$ be the reference ground truth ENF and $\tilde{\mathbf{g}}(L)=[g_l,g_{l+1},\ldots,g_{l+K-1}]^\top$ be a segment of g starting at l. The following index is determined:

$$l_{opt}=\arg\max_{l}c(l) \tag{13}$$

where $l=1,2,\ldots,\tilde{K}-K+1$ and $c(l)$ is the sample correlation coefficient between $\mathbf{f}$ and $\tilde{\mathbf{g}}(L)$ defined as:

$$c(l)=\frac{\mathbf{f}^\top\tilde{\mathbf{g}}(l)}{\|\mathbf{f}\|_2\|\tilde{\mathbf{g}}(l)\|_2}. \tag{14}$$

For the second metric, mSDE, the best index is found by minimising the squared error between $\mathbf{f}$ and $\tilde{\mathbf{g}}(L)$, that is,

$$l_{opt}=\underset{l=1}{\overset{\tilde{K}-K+1}{\mathrm{argmin}}}\|\mathbf{f}-\tilde{\mathbf{g}}(l)\|_2^2. \tag{15}$$

In Section 4, pairwise differences between the MCC delivered by the proposed approach and that of state-of-the-art ENF estimations were calculated in order to assess whether they are statistically significant. Fisher's transformation was employed for this purpose. A similar procedure was followed regarding the statistical significance of the differences delivered by mSDE. Dynamic time warping can also be used as an alternative for matching the extracted ENF to the ground truth [36].

4 | RESULTS AND DISCUSSION

The approach proposed in Section 3 was applied to the two data sets described in Section 2.2. The ENF was estimated every second for the total duration of 30 min in each data set employing the parameters shown in Table 2. Regarding Data 1, the first, second, and third harmonics were used for ENF estimation. Regarding Data 2, only the second harmonic was

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

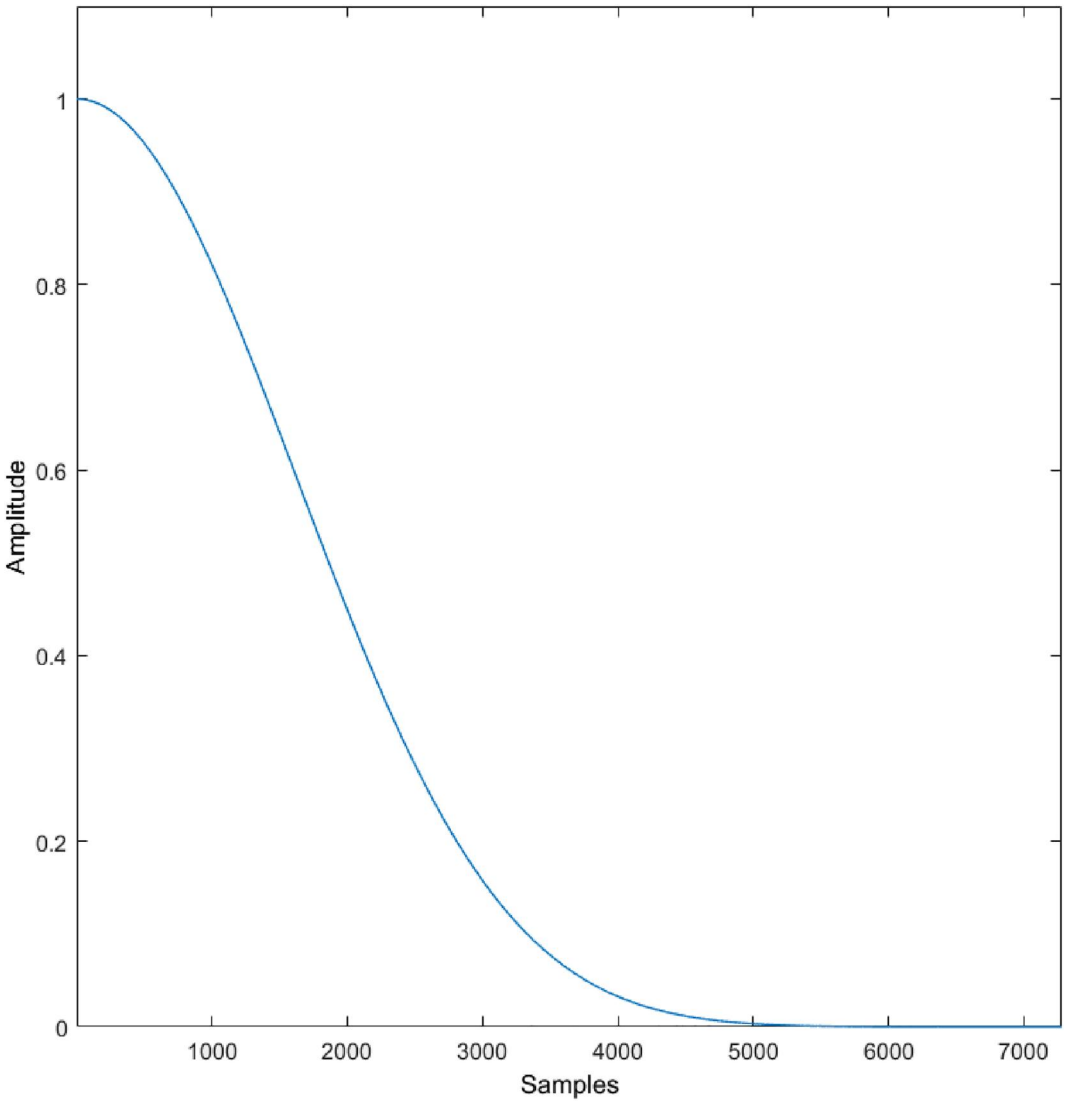

FIGURE 1 Window function $w(\zeta)$ in the time domain for $M = 0.5LF_s$ (Data 2)

exploited, because the other two harmonics were too weak. This was due to the strong interference present in speech recordings. In order to evaluate the proposed approach, comparisons were made against state-of-the-art methods, applied to the same data sets using the same setup.

4.1 | Data 1

The high SNR nature of Data 1 enables accurate ENF estimation for all three harmonics. For the first harmonic of ENF, the proposed BT + LWD, which employs the designed lag window resulted to MCC of 0.9990 outperforming the Welch method, which resulted in an MCC of 0.9983 in [36]. There, the MCC between the ENF signal estimated by the BT method and the ground truth ENF was measured to be 0.9924. The lag window employed in the aforementioned case of BT estimate was the rectangular one. An MCC of 0.9900 was reported in [28] for the Time-Recursive Iterative Adaptive Approach with frequency tracking, that is, TRIAA (Track), while their proposed approach TRIAA without Track, reached 0.9895. The same figure of merit was 0.9917 for the STFT (Track) implementation [28]. The fast version of Capon estimator, which combines the quality of power spectral estimate delivered by the Capon method with low computational time requirements, achieves an MCC of 0.9922 [40]. The proposed BT + LWD outperformed all state-of-the-art approaches employing a value of $M = 0.25LF_s$ and $\theta = 3.5/M$ for the designed lag window. This approach constitutes a reliable tool for ENF estimation, because of the high accuracy and low time requirements. Specifically, it outperforms the state-of-the-art TRIAA (Track) both in terms of correlation coefficient and computational complexity. Generally, periodogram-based methods perform better in the presence of high SNR. Regarding the second harmonic, there seemed to be more interference than with the first one. The proposed approach BT + LWD resulted in an MCC of 0.9930. BT + LWD outperforms TRIAA method, which resulted in an MCC of

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

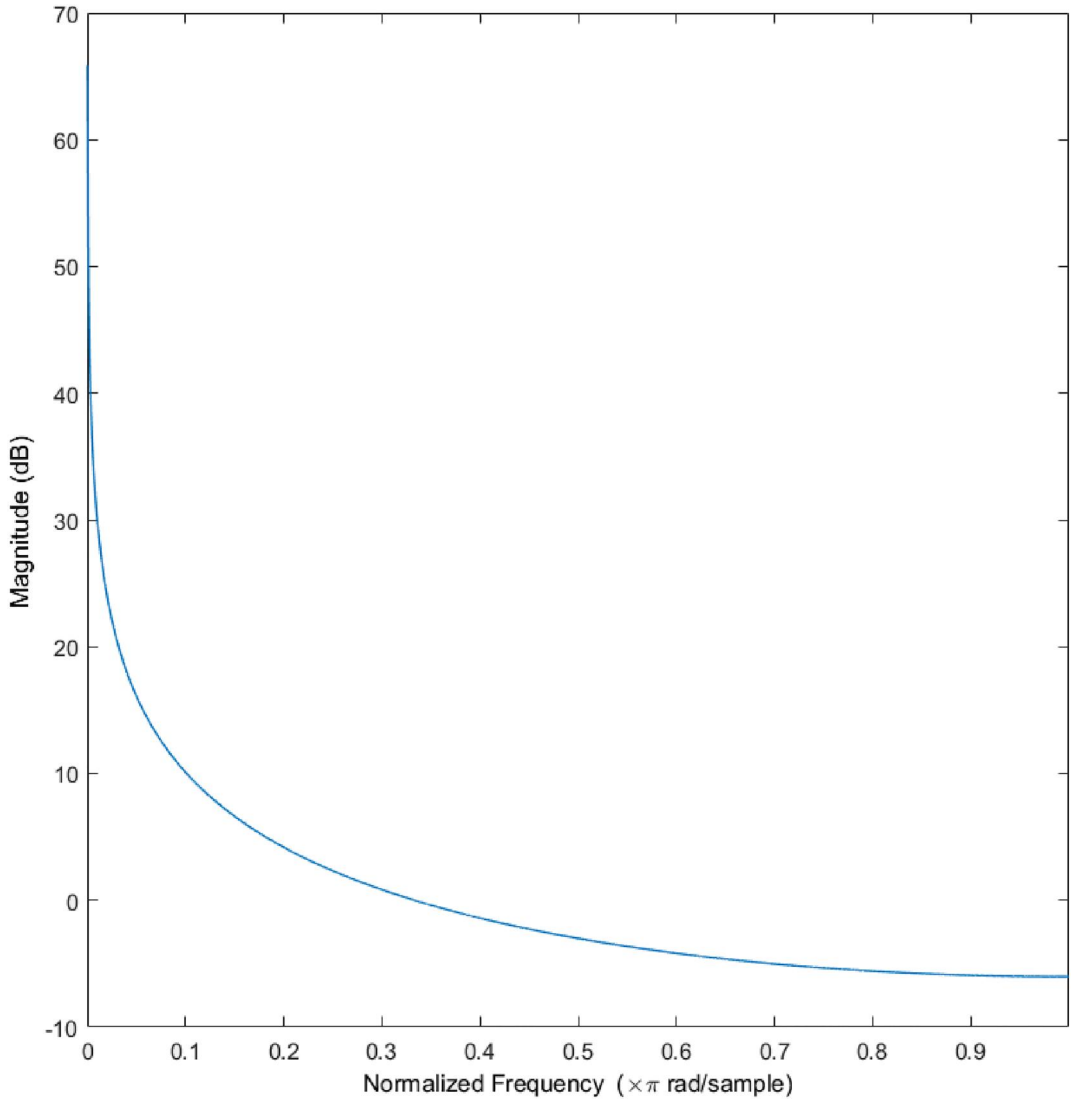

FIGURE 2 Frequency response of spectral window $W(\omega)$ for $M = 0.5LF_s$ (Data 2)

0.9902. Fast Capon also performed well and achieved an MCC of 0.9913. The proposed approach outperforms Welch and BT methods, which resulted in an MCC of 0.985 for both. The third harmonic is the most prominent. A large number of results disclosed in the literature for ENF estimation refer to the third harmonic, exclusively. The proposed BT + LWD achieved an MCC of 0.9990 for $M = 0.25LF_s$ and $\theta = 3.5/M$, outperforming all its competitors in ENF estimation at the third harmonic. The ML approach [29] yielded an MCC of 0.9977, while the conventional BT approach delivered an MCC of 0.9978 in [36]. MCC was 0.9961 for the TRIAA (Track), lagging behind STFT (Track). When employing the linear prediction approach [35], MCC was measured at 0.9982, while its enhanced denoising version, called linear prediction with robust principal component analysis (RPCA), reached an MCC of 0.9984, the top performance at the time it was published. Furthermore, remarkable results were obtained by fast Capon, yielding an MCC of 0.9990. The proposed approach improved ENF estimation, outperforming its competitors and it was established as a reliable solutions for all harmonics. At this point, it is worth mentioning that ENF estimation in a single harmonic is considered to be sufficient enough in real-world applications. Many works in the literature estimate ENF employing one harmonic only. Detailed results for all harmonics and approaches are shown in Table 3.

The differences between the proposed approach and all other methods should be assessed as to whether they are statistically significant. To this end, hypothesis testing was applied, employing Fisher transformation. The null hypothesis, $H_0 : c_1 = c_2$, indicates that MCCs are equal and the alternative one, H_1: $c_1 \neq c_2$, indicates the opposite. For each pair of approaches, the associated MCCs undergo Fisher's z transformation [64]:

$$z = 0.5 \ln \frac{1 + c}{1 - c}. \tag{16}$$

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Maximum correlation coefficient for various approaches applied to Data 1

Approach	60 Hz	120 Hz	180 Hz
Linear prediction [35]	–	–	0.9982
Linear prediction (RPCA) [35]	–	–	0.9984
ML [29]	–	–	0.9977
Welch [36]	0.9983	0.9850	0.9983
BT [36]	0.9924	0.9850	0.9978
BT + LWD	0.9990	0.9930	0.9990
TRIAA [28]	0.9895	0.9902	0.9961
TRIAA (Track) [28]	0.9900	0.9946	0.9961
STFT [28]	0.9912	0.9911	0.9968
STFT (Track) [28]	0.9917	0.9949	0.9968
Fast Capon [40]	0.9922	0.9913	0.9990

The number of samples is $K = 1800$. The test statistic is given by:

$$q_F = \sqrt{K-3}\,(z_1 - z_2) \tag{17}$$

Comparisons were conducted between the proposed approach at each harmonic and the rest of the approaches with respect to MCC, as reported in Table 3. In all cases, q_F was calculated and found to be outside the region of acceptance for a significance level of 5%. Consequently, there is sufficient evidence to warrant the rejection of the null hypothesis at a significance level of 5%. Moreover, hypothesis testing at a significance level of 1% was conducted. In all cases, q_F was calculated and found to be outside the region of acceptance $-2.58 < q_F < 2.58$. Consequently, there is sufficient evidence to warrant the rejection of the null hypothesis at a significance level of 1%, too. Therefore, the differences between MCC are statistically significant. It is worth mentioning that the proposed approach demonstrates statistically significant improvements compared to the conventional BT method. Moreover, an L-size Parzen window was employed as a temporal window prior to spectral estimation, as explained in Section 2.2.

The second metric employed is the mSDE. The empirical findings are summarised in Table 4. A limited amount of papers employ this metric. It is worth mentioning that there are approaches, which although they yield MCC exceeding 0.99, their mSDE is relatively high. For this reason, experiments employing mSDE are conducted systematically to point out the efficacy of the approach under examination.

As stated previously, better results are obtained when the third harmonic is used. BT + LWD mSDE at the first harmonic was measured $0.838 \cdot 10^{-3}$, outperforming STFT (Track) and TRIAA (Track), which yielded mSDEs of $2.650 \cdot 10^{-3}$ and $2.919 \cdot 10^{-3}$, respectively. Moreover, the proposed BT + LWD outperforms BT, which resulted in an mSDE of $2.284 \cdot 10^{-3}$. TRIAA reached an mSDE of $3.032 \cdot 10^{-3}$, while

TABLE 4 Minimum standard deviation of error for approaches applied to Data 1

Approach	60 Hz	120 Hz	180 Hz
STFT [28]	2.772×10^{-3}	2.774×10^{-3}	1.900×10^{-3}
STFT (Track) [28]	2.650×10^{-3}	2.145×10^{-3}	1.851×10^{-3}
ML [29]	–	–	0.760×10^{-3}
BT [36]	2.284×10^{-3}	3.202×10^{-3}	1.218×10^{-3}
BT + LWD	0.838×10^{-3}	1.746×10^{-3}	0.832×10^{-3}
TRIAA [28]	3.032×10^{-3}	2.822×10^{-3}	1.999×10^{-3}
TRIAA (Track) [28]	2.919×10^{-3}	2.198×10^{-3}	1.999×10^{-3}

conventional STFT attained $2.772 \cdot 10^{-3}$. The proposed BT + LWD employed $M = 0.25LF_s$ and $\theta = 3.5/M$. Regarding the second harmonic, the proposed BT + LWD achieved an mSDE of 1.746×10^{-3}, outperforming all its competitors. STFT (Track) yielded an mSDE value of 2.145×10^{-3}, while TRIAA (Track) resulted in $2.198 \cdot 10^{-3}$. The results deteriorated when TRIAA was employed. An mSDE value of 2.822×10^{-3} was reached, lagging behind STFT, which achieved a value of 2.774×10^{-3}. BT + LWD mSDE at the third harmonic was measured 0.832×10^{-3}, outperforming TRIAA (Track) mSDE, which was measured at 1.999×10^{-3} [28]. The conventional STFT yielded an mSDE of 1.900×10^{-3} for the third harmonic, while an increase was noticed at the first and second harmonics, reaching 2.772×10^{-3} and 2.774×10^{-3}, respectively. The STFT (Track) approach reached a value of 1.851×10^{-3}, which was slightly better than the conventional one. BT reached a value of 1.218×10^{-3}, also lagging behind the proposed BT + LWD. mSDE of the ML approach reached a value of $0.760 \times` 10^{-3}$. In order to determine whether the differences between mSDE of the top-performing approaches and the rest of the approaches reported in Table 4 are statistically significant, hypothesis testing was applied [65]. Under the assumption that the errors are normally distributed, the statistical test for the variances of errors was based on the F-distribution. The sample variances s^2 of errors for each method were calculated. The null hypothesis H_0: $s_1^2 = s_2^2$ indicates that the variances of errors are equal, while the alternative hypothesis indicates the opposite. For a significance level of 5% and number of samples $K = 1800$, a two-tailed test was applied. The test statistic for each pair of approaches is [65]:

$$q_t = \frac{s_1^2}{s_2^2}. \tag{18}$$

Comparisons were conducted between the top-performing approach at each harmonic and the rest of the approaches with respect to the mSDE, as reported in Table 4. For all pairs of comparisons, q_t was found to be outside the region of acceptance $0.9116 \leq q_t \leq 1.0968$. Thus, there is sufficient evidence to warrant the rejection of the null hypothesis at a significance level of 5%. Accordingly, the differences between the variances of errors are statistically significant. Furthermore,

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

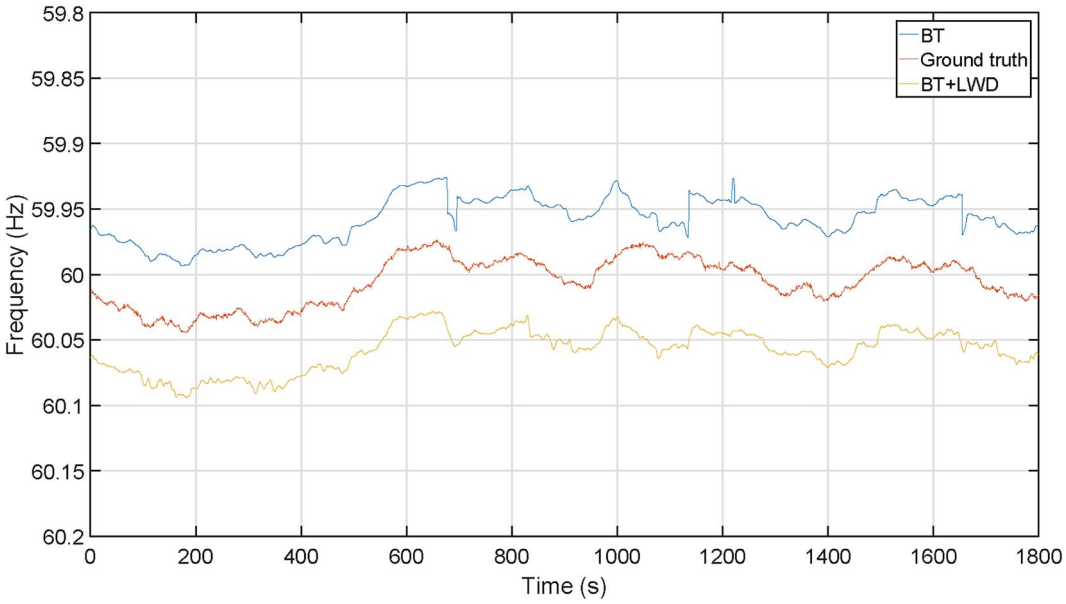

FIGURE 3 Extracted ENF signal using BT and BT + LWD employing the proposed lag window to Data 2

hypothesis testing at a significance level of 1% was performed. In all cases, q_t was calculated and found to be outside the region of acceptance $0.89 < q_t < 1.13$. Consequently, there is sufficient evidence to warrant the rejection of the null hypothesis at a significance level of 1% for all comparisons and state that the variances of errors are statistically significant. The proposed approach BT + LWD outperforms in terms of mSDE the TRIAA (Track) and demonstrates statistically significant improvements compared to the conventional BT method. The state-of-the-art methods proposed in [35] did not provide results regarding mSDE. To conclude, BT + LWD was found to be the top-performing approach for ENF estimation in the first and second harmonics and the second best approach for ENF estimation in the third harmonic with respect to mSDE.

4.2 | Data 2

An example of an extracted ENF signal using the conventional BT and BT employing the proposed designed lag window, i.e. BT + LWD, is shown in Figure 3. The reference ground truth signal is also overlaid in Figure 3. The ENF signals are shifted by 0.05 Hz up or down from their actual values for illustration purposes, in Figure 3. As can be noticed, at around 700 s, the proposed BT + LWD provides more accurate extraction of ENF compared to the conventional BT. The same is indicated around 1100, 1250, and 1650 s. These improvements can also be detected by visual inspection of absolute errors between each approach and the reference ground truth. These differences tended to zero when BT + LWD was employed, as can be seen in Figure 4.

Regarding Data 2, only the second harmonic was exploited, as was done in the literature, due to the fact that the other two harmonics were too weak. Data 2 is more challenging than Data 1, being closer to real-world practical forensic applications. The difficulty in ENF estimation lies in the fact that the pitch frequency for a male (135 Hz for vowel/a/) or female voice (200 Hz for vowel/a/) in Data 2 interferes with ENF harmonics. The variability of pitch is large both for females and males and, thus, has a serious impact on ENF estimation in Data 2. Specifically, the range of pitch frequencies is approximately 120–350 Hz and 100–200 Hz for females and males, respectively [66]. The proposed BT + LWD delivered an MCC value of 0.9434, outperforming the recent state-of-the-art approach in ENF estimation, that is, linear prediction, [35], which yields an MCC of 0.9366. Moreover, the proposed approach also outperforms TRIAA, which yields an MCC value of 0.9305. The conventional STFT, which is often adopted in the literature, resulted in an MCC of 0.9125. The fast implementation of the Capon method [40] yielded an MCC of 0.9351. The ML approach [29] reached an MCC of 0.9319, outperforming the Welch method [36] that reached 0.9179. It is worth pointing out that the best performing approaches in Data 1 turn out to be fragile in Data 2. The MCC of BT estimator was measured at 0.9179, lagging behind the proposed BT + LWD. Linear prediction (RPCA) and TRIAA (Track) seem to perform better than the proposed BT + LWD, but there is a key difference that should be noted. The approaches that outperform the proposed BT + LWD include an extra module, namely, RPCA or Track. As a result, the proposed method facilitates generalisation and can be employed in a variety of applications and spectral analysis frameworks. Furthermore, there is a significant difference between the proposed BT + LWD and TRIAA. The proposed BT + LWD computational complexity is about $\Xi\left[\frac{1}{2}\log_2(\Xi) + 2\log_2(2\Xi)\right]$, while the bottleneck of the TRIAA approach lies in its high complexity of about $\mathcal{O}(N^2\Xi)$. Employing $\Xi = 4N$, where

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

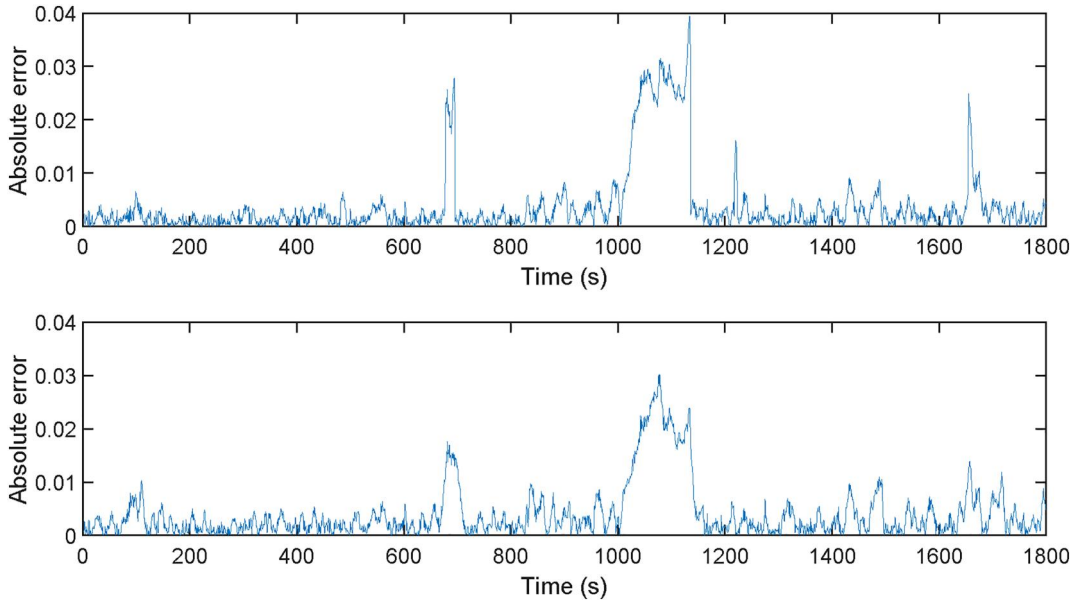

FIGURE 4 Absolute errors between the reference ground truth and the extracted ENF for Data 2, employing (a) BT and (b) BT + LWD

N is the segment length, $\Xi\left[\frac{1}{2}\log_2(\Xi)+2\log_2(2\Xi)\right]=4N\left[\log_2(N)+6\log_2(N)\right]=28N\log_2 N$. In the same manner, TRIAA computational complexity results to $\mathcal{O}(4N^3)$.

In order to design an effective lag window able to provide optimal results in terms of leakage effects, two critical factors should be taken into consideration. Both factors affect crucially the outcome of the spectral analysis. The first factor is related to lag window length in samples, M, which should be chosen carefully in order to achieve an optimal trade-off between statistical variance and spectral resolution. The second factor that should be taken into consideration is parameter θ. Thus, the proper choice of θ should offer an optimal trade-off between the energy in the main lobe and the side lobe. There is not any generally applicable rule of thumb. It is worth mentioning that the design parameter θ should be larger than $1/M$. Otherwise, spectral window design will deteriorate the results and leakage reduction will be hindered. The authors conducted a systematic study to choose the optimal values for both parameter θ and lag window length M, as shown in Figure 5. The curves depict the MCC for various values of the aforementioned parameters. The optimal value of MCC for the challenging Data 2 is obtained for $\theta = 2.1/M$ and $M = 0.5LF_s$. MCCs for various approaches applied to the second harmonic of Data 2 are summarised in Table 5.

As discussed in Section 4.1, in order to assess whether MCC differences for any pair of approaches are statistically significant, Fisher's transformation was applied. In all cases, q_F was calculated and found to be outside the region of acceptance for a significance level of 5%. Consequently, there is sufficient evidence to warrant the rejection of the null hypothesis at a significance level of 5%. Furthermore, hypothesis testing at a significance level of 1% was conducted. In all cases, q_F was calculated and found to be outside the region of acceptance. Consequently, there is sufficient evidence to warrant the rejection of the null hypothesis at a significance level of 1%, too. Accordingly, BT + LWD is ranked as the fourth best approach, behind approaches employing an additional module, such as TRIAA (Track), STFT (Track), and linear prediction (RPCA). BT + LWD delivers a statistically significant difference against linear prediction, fast Capon, ML, TRIAA, BT, and Welch.

Regarding the mSDE, the proposed BT + LWD delivered an mSDE of 6.287×10^{-3}, outperforming TRIAA, which yielded a value of 7.225×10^{-3}. The conventional STFT yielded a value of 7.948×10^{-3}. BT spectral estimation deliver accurate results, reaching an mSDE of 7.052×10^{-3}. Moreover, the Welch method achieved an mSDE of 7.313×10^{-3}. Generally, strong interference in this data set hinders accurate ENF estimation. TRIAA (Track), which is based on dynamic programming, reached an mSDE of 2.914×10^{-3} and the ML approach yielded a value of 3.839×10^{-3} [29]. Moreover, STFT (Track) achieved an mSDE value of 3.369×10^{-3}. The results of a linear prediction approach regarding mSDE were not disclosed in [35]. The mSDE measured for various approaches applied to Data 2 is listed in Table 6.

Statistical tests were conducted to assess whether mSDE differences for any pair of approaches are statistically significant at significance levels of 5% and 1%. In both sets of tests, there is sufficient evidence to warrant the rejection of the null hypothesis at the aforementioned significance levels. Accordingly, BT + LWD is ranked as the fourth best approach behind TRIAA (Track), ML, and STFT (Track). BT + LWD delivers a statistically significant mSDE difference against BT, TRIAA, Welch, and STFT.

5 | CONCLUSIONS

Strong interference present in speech recordings hinders ENF estimation. An approach based on the design of a lag window for a BT estimator has been proposed. Such a lag window

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

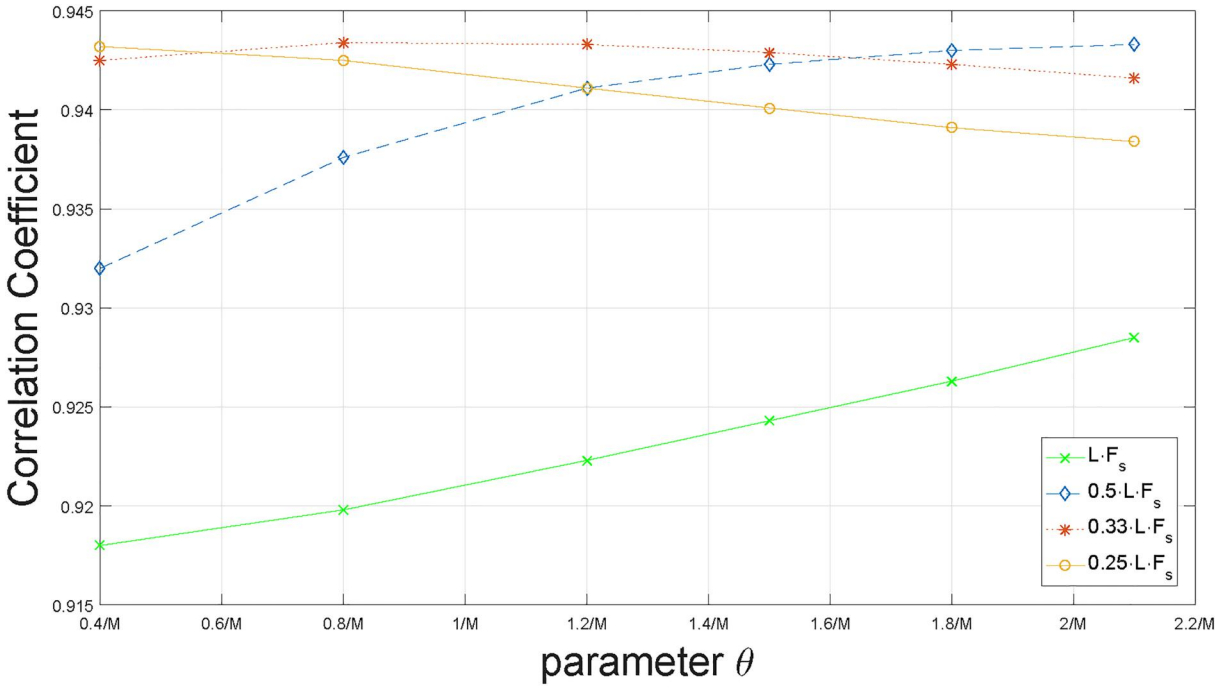

FIGURE 5 Correlation coefficient of the proposed BT + LWD approach for various values of the design parameter θ and lag window length M in Data 2

TABLE 5 Maximum correlation coefficient for various approaches applied to Data 2

Approach	120 Hz
Linear prediction [35]	0.9366
Linear prediction (RPCA) [35]	0.9764
ML [29]	0.9319
Welch [36]	0.9179
BT [36]	0.9179
BT + LWD	0.9434
TRIAA [28]	0.9305
TRIAA (Track) [28]	0.9907
STFT [28]	0.9125
STFT (Track) [28]	0.9857
Fast Capon [40]	0.9351

TABLE 6 Minimum standard deviation of error for approaches applied to Data 2

Approach	120 Hz
STFT [28]	7.948×10^{-3}
STFT (Track) [28]	3.369×10^{-3}
ML [29]	3.839×10^{-3}
Welch	7.313×10^{-3}
BT	7.052×10^{-3}
BT + LWD	6.287×10^{-3}
TRIAA [28]	7.225×10^{-3}
TRIAA (Track) [28]	2.914×10^{-3}

design delivers an optimal trade-off between smearing and leakage and guarantees a reliable ENF estimation under various SNR conditions. The proposed approach has been tested on real-world data sets and has been compared to state-of-the-art ENF estimation methods. It has been demonstrated that it outperforms many competitors, revealing its potential in ENF estimation. Statistical tests have attested that performance improvements are statistically significant. Testing the proposed approach in more challenging contexts, including ENF extraction from images and videos, could be a topic of future research.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. D. Bykhovsky and Dr. A. Cohen for sharing their algorithms. This research has been financially supported by the General Secretariat for Research and Technology (GSRT) and the Hellenic Foundation for Research and Innovation (HFRI) (Scholarship Code: 820).

ORCID

Constantine Kotropoulos https://orcid.org/0000-0001-9939-7930

REFERENCES

1. Barbara, J.J.: Handbook of Digital and Multimedia Forensic Evidence. Springer Science & Business Media (2007)
2. Popescu, A.C., Farid, H.: Statistical tools for digital forensics. In: Proc. Int. Work. Inf. Hiding, Berlin, pp. 128–147. (2004)

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. Ho, A.T.S., Li, S.: Handbook of Digital Forensics of Multimedia Data and Devices. John Wiley & Sons (2015)
4. Manral, B., et al.: A systematic survey on cloud forensics challenges, solutions, and future directions. ACM Comput. Surv. 52(6), pp. 1–38. (2019)
5. Roussev, V., et al.: Cloud forensics—Tool development studies & future outlook. Digit Invest. 18, pp. 79–95. (2016)
6. Farid, H.: Photo Forensics. MIT Press (2016)
7. Fridrich, J.: Digital image forensics. IEEE Signal Process Mag. 26(2), pp. 26–37. (2009)
8. Piva, A.: An overview on image forensics. ISRN Signal Process. 2013 (2013)
9. Krawetz, N.: Digital photo forensics. Handbook of Digital Imaging, pp. 1–34 (2015). https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118798706.hdi044
10. Wang, X., et al.: Keypoints-based image passive forensics method for copy-move attacks. Int. J. Pattern Recogn. Artif. Intell. 30(03), 1655008 (2016)
11. Amerini, I., et al.: A SIFT-based forensic method for copy-move attack detection and transformation recovery. IEEE Trans. Inform. Forensic Secur. 6(3), pp. 1099–1110. (2011)
12. Ardizzone, E., Bruno, A., Mazzola, G.: Copy-move forgery detection by matching Triangles of Keypoints. IEEE Trans. Inform. Forensic Secur. 10(10), pp. 2084–2094. (2015)
13. Zhao, J., Guo, J.: Passive forensics for copy-move image forgery using a method based on DCT and SVD. Forensic Sci. Int. 233(1–3), pp. 158–166. (2013)
14. Bayram, S., Sencar, H.T., Memon, N.: An efficient and robust method for detecting copy-move forgery. In: Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process, Taipei, pp. 1053–1056. (2009)
15. Chang, C.C., Hu, Y.S., Lu, T.C.: A watermarking-based image ownership and tampering authentication scheme. Pattern Recogn. Lett. 27(5), pp. 439–446. (2006)
16. Dadkhah, S., et al.: An effective SVD-based image tampering detection and self-recovery using active watermarking. Signal Process Image Commun. 29(10), pp. 1197–1210. (2014)
17. Qin, C., et al.: Non-uniform watermark sharing based on optimal iterative BTC for image tampering recovery. IEEE Multimedia. 25(3), pp. 36–48. (2018)
18. Shehab, A., et al.: Secure and robust fragile watermarking scheme for medical images. IEEE Access. 6, pp. 10269–10278. (2018)
19. Maher, R.: Audio forensic examination. IEEE Signal Process Mag. 26(2), pp. 84–94. (2009)
20. Wang, S., et al.: Detection of speech tampering using sparse representations and spectral manipulations based information hiding. Speech Commun. 112, pp. 1–14. (2019)
21. Huang, X., et al.: Fast and effective copy-move detection of digital audio based on auto segment. Int. J. Digital Crime Forensics (IJDCF). 11(2), pp. 47–62. (2019)
22. Rumsey, F.: Audio forensics: Keeping up in the age of smartphones and fakery. J. Audio Eng. Soc. 67(7/8), pp. 617–622. (2019)
23. Malik, H., Farid, H.: Audio forensics from acoustic reverberation. In: Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process, Dallas, TX, USA, pp. 1710–1713. (2010)
24. Ikram, S., Malik, H.: Digital audio forensics using background noise. In: Proc. IEEE Int. Conf. Multimedia and Expo, Suntec City, Singapore, pp. 106–110. (2010)
25. Jadhav, S., Patole, R., Rege, P.: Audio splicing detection using convolutional neural network. In: Proc. 10th Int. Conf. Computing, Communication and Networking Technologies, Kanpur, India, pp. 1–5. (2019)
26. Grigoras, C.: Digital audio recording analysis: the electric network frequency (ENF) criterion. Int. J. Speech Lang. Law. 12(1), pp. 63–76. (2005)
27. Grigoras, C.: Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis. Forensic Sci. Int. 167(2–3), pp. 136–145. (2007)
28. Ojowu, O. et al.: ENF extraction from digital recordings using adaptive techniques and frequency tracking. IEEE Trans. Inform. Forensic Secur. 7(4), pp. 1330–1338. (2012)
29. Bykhovsky, D., Cohen, A.: Electrical network frequency (ENF) maximum-likelihood estimation via a multitone harmonic model. IEEE Trans. Inf. Forensics Secur. 8(5), 774–753 (2013)
30. Hajj-Ahmad, A., Garg, R., Min Wu, M.: Spectrum combining for ENF signal estimation. IEEE Signal Process Lett. 20(9), pp. 885–888. (2013)
31. Fu, L., et al.: An improved discrete Fourier transform-based algorithm for electric network frequency extraction. IEEE Trans. Inf. Forensics Secur. 8(7), pp. 1173–1181. (2013)
32. Kajstura, M., Trawinska, A., Hebenstreit, J.: Application of the electrical network frequency (ENF) criterion: a case of a digital recording. Forensic Science Int. 155(2-3), pp. 165–171. (2005)
33. Huijbregtse, M., Geradts, Z.: Using the ENF criterion for determining the time of recording of short digital audio recordings. In: Proc. Int. Work. Computational Forensics, The Hague, pp. 116–124. (2009)
34. Cooper, A.J.: An automated approach to the electric network frequency (ENF) criterion – theory and practice. Int. J. Speech Lang. Law. 16(2) (2009)
35. Lin, X., Kang, X.: Robust electric network frequency estimation with rank reduction and linear prediction. ACM Trans. Multimedia Comput. Commun. Appl. 14(4), pp. 1–13. (2018)
36. Karantaidis, G., Kotropoulos, C.: Assessing spectral estimation methods for electric network frequency extraction. In: Proc. 22nd Panhellenic Conf. Informatics, Athens, Greece, pp. 202–207. (2018)
37. Dosiek, L.: Extracting electrical network frequency from digital recordings using frequency demodulation. IEEE Signal Process Lett. 22(6), pp. 691–695. (2015)
38. Hajj-Ahmad, A., Garg, R., Wu, M.: Instantaneous frequency estimation and localization for ENF signals. In: Proc. Asia Pacific Signal and Inf. Process. Assoc. Annual Summit and Conference, Hollywood, pp. 1–10. (2012)
39. Triantafyllopoulos, A., et al.: A Hilbert-based approach to the ENF extraction problem. Proc. IEICE Inf. and Communication Technology Forum, Patras, Greece (2016)
40. Karantaidis, G., Kotropoulos, C.: Efficient Capon-based approach exploiting temporal windowing for electric network frequency estimation. Proc. IEEE 29th Int. Work. Machine Learning for Signal Process
41. Hua, G., Zhang, H.: ENF signal enhancement in audio recordings. IEEE Trans. Inform. Forensic Secur. 15, pp. 1868–1878. (2020)
42. Elmesalawy, M.M., Eissa, M.M.: New forensic ENF reference database for media recording authentication based on harmony search technique using GIS and wide area frequency measurements. IEEE Trans. Inform. Forensic Secur. 9(4), pp. 633–644. (2014)
43. Yao, W., et al.: Source location identification of distribution-level electric network frequency signals at multiple geographic scales. IEEE Access. 5, pp. 11166–11175. (2017)
44. Hajj-Ahmad, A., Garg, R., Wu, M.: ENF-based region-of-recording identification for media signals. IEEE Trans. Inform. Forensic Secur. 10(6), pp. 1125–1136. (2015)
45. Lin, X., Liu, J., Kang, X.: Audio recapture detection with convolutional neural networks. IEEE Trans. Multimedia. 18(8), 1480–1487 (2016)
46. Hajj-Ahmad, A., et al.: Factors affecting ENF capture in audio. IEEE Trans. Inform. Forensic Secur. 14(2), pp. 277–288. (2019)
47. Bang, W., Yoon, J.W.: Forecasting the electric network frequency signals on power grid. Proc. Int. Conf. Inf. and Communication Technology Convergence, Jeju-si, Jeju-do, S. Korea, pp. 1218–1223. (2019)
48. Esquef, P.A.A., Apolinario, J.A., Biscainho, L.W.P.: Edit detection in speech recordings via instantaneous electric network frequency variations. IEEE Trans. Inform. Forensic Secur. 9(12), pp. 2314–2326. (2014)
49. Rodriguez, D.P.N., Apolinario, J.A., Biscainho, L.W.P.: Audio authenticity: detecting ENF discontinuity with high precision phase analysis. IEEE Trans. Inf. Forensics Secur. 5(3), pp. 534–543, (2010)
50. Reis, P.M.J.I., et al.: Audio authentication using the kurtosis of ESPRIT based ENF estimates. Proc. 10th Int. Conf. Signal Process. and Communication Systems, Gold Coast, Australia (2016)
51. Reis, P.M.G.I., et al.: ESPRIT-Hilbert-based audio tampering detection with SVM classifier for forensic analysis via electrical network frequency. IEEE Trans. Inform. Forensic Secur. 12(4), pp. 853–864. (2017)

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

52. Wang, Z., et al.: Digital audio tampering detection based on ENF consistency. In: Proc. Int. Conf. Wavelet Analysis and Pattern Recognition, Sichuan, China, pp. 209–214. (2018)
53. Hu, Y., et al.: Audio forgery detection based on max offsets for cross correlation between ENF and reference signal. In: Proc. Int. Work. Digital Forensics and Watermarking, Auckland, pp. 253–266. (2013)
54. Wang, Y., et al.: ENF based video forgery detection algorithm. Int. J. Digit. Crime Forensics. 12(1), pp. 131–156. (2020)
55. Hua, G., et al.: Audio authentication by exploring the absolute-error-map of ENF signals. IEEE Trans. Inform. Forensic Secur. 11(5), pp. 1003–1016, (2016)
56. Su, H., et al.: Exploring the use of ENF for multimedia synchronization. In: Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process, Florence, pp. 4613–4617. (2014)
57. Nicolalde-Rodríguez, D.P., Apolinário, J.A., Biscainho, L.W.P.: Audio authenticity based on the discontinuity of ENF higher harmonics. In: Proc. 21st European Signal Process. Conference, Marrakech, Morocco, pp. 1–5. (2013)
58. ENTSOE: P1—Policy 1: Load-frequency control and performance. In: Operations Handbook (2009). www.entsoe.eu/fileadmin/userupload/library/publications/ce/oh/Policy1final.pdf
59. Parzen, E.: Mathematical considerations in the estimation of spectra. Technometrics. 3(2), pp. 167–190. (1961)
60. Harris, F.J.: On the use of windows for harmonic analysis with the discrete Fourier transform. Proc. IEEE. 66(1), pp. 51–83. (1978)
61. Smith, J.O., Serra, X.: PARSHL: An analysis/synthesis programme for non-harmonic sounds based on a sinusoidal representation. In: Proc. Int. Computer Music Conference, Champaign/Urbana, IL, USA, pp. 290–297. (1987)
62. Stoica, P., Moses, R.L.: Spectral Analysis of Signals. Pearson Prentice Hall, Upper Saddle River (2005)
63. Cooper, A.J.: The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings–an automated approach. Proc. Audio Engineering Society 33rd Int. Conf. Audio Forensics-Theory and Practice
64. Papoulis, A.: Probability and Statistics. Prentice-Hall, Inc. (1990)
65. Devore, J.L., Berk, K. N.: Modern Mathematical Statistics with Applications. Springer (2012)
66. Titze, I.R., Martin, D.W.: Principles of Voice Production. Acoustical Society of America (1998)

How to cite this article: Karantaidis, G., Kotropoulos, C.: Blackman–Tukey spectral estimation and electric network frequency matching from power mains and speech recordings. IET Signal Process. 15(6), 396–409 (2021). https://doi.org/10.1049/sil2.12039

17519683, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/sil2.12039 by Cochrane Greece, Wiley Online Library on [20/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License