Available online at www.sciencedirect.com
ScienceDirect
Procedia Technology 21 (2015) 596 – 603
Procedia Technology

SMART GRID Technologies, August 6-8, 2015

Audio Data Authentication with PMU Data and EWT

Premjith B[a]*, Neethu Mohan[a], Prabaharan Poornachandran[b], Soman K P[a]

[a]Centre for Excellence in Computational Engineering and Networking, Amrita Vishwa Vidyapeetha,, Coimbatore – 641112, India
[b]Amrita Center for Cyber Security Systems and Networks, Amrita Vishwa Vidyapeetham, Kollam

Abstract

Digital forensics has become a flourishing research area. Electrical Network Frequency (ENF) plays an important role in assessing the authenticity of a digital recording such as audio. ENF criterion is a tool for extracting the embedded power line frequency from the recording. A cross correlation between a reference PMU data and extracted ENF signal can be done in order to determine the authenticity of an audio signal. In this paper, Empirical Wavelet Transform (EWT) is used for extracting the ENF from an audio signal. EWT decomposes signal into *N* modes. Hilbert Transform is used to compute the instantaneous frequency and amplitude of the extracted mode corresponding to ENF. EWT method is not able to capture the weak harmonics in a signal. This problem is resolved by fixing the frequency domain boundaries of each mode.

Peer-review under responsibility of Amrita School of Engineering, Amrita Vishwa Vidyapeetham University

Keywords: Digital Forensics; Electrical Network Frequency; Empirical Wavelet Transform; Hilbert Transform

1. Introduction

Digital recordings are increasingly used as a critical evidence in legal proceedings. Ensuring the authenticity of a digital recording is very difficult compared to analogue signals. Electrical Network Frequency (ENF) criterion proposed by Grigoras [1] is a tool for assessing the authenticity of an audio signal by extracting the ENF from an audio signal and comparing it with the ENF obtained from PMU data taken on the same time. ENF criterion also allows us to identify the time and location of the recording. ENF is approximately a constant voltage frequency gets added to all the signals recorded with a device connected to a power line. ENF for a network is fixed and is 50Hz in India and U.K and 60Hz in USA. Usually an ENF signal contains fundamental frequency (50Hz/60Hz) and its odd harmonics.

Grigoras et.al [1], [2] and M. Kajstura et.al [3] describes the application of ENF in audio and video authentications. ENF extraction procedures are mainly categorized into three.

*Corresponding author. Tel.: +919597141816;
*E-mail address:*prem.jb@gmail.com

Peer-review under responsibility of Amrita School of Engineering, Amrita Vishwa Vidyapeetham University
doi:10.1016/j.protcy.2015.10.066

- Time – Frequency domain analysis: In this approach, spectrogram of a signal is compared visually with a reference data.
- Frequency domain analysis: A band pass filter is applied to extract the desired frequency location.
- Time domain analysis: A band pass filter is applied first and then measure the zero crossing rate.

Many approaches have been adopted for audio authentication. D. Varodayan et.al [4] proposed a distributed source coding based audio authentication scheme. They provided users with a Slepian – Wolf encoded audio projection which can be decoded at client side. Farid [5] proposed a bispectral analysis for assessing the integrity of an audio signal. Yang et.al [6] detected audio forgeries by checking frame offsets. M. Kajstura et.al [3], Yuming Liu et.al [7], Ode Ojowu Jr. wt.al [8] and Grigoras [17] used ENF criterion for audio authentication. Yuming Liu et.al [7] used STFT (Short Term Fourier Transform) for extracting ENF while Ode Ojowu Jr et.al [8] used some data dependent filters for taking out ENF from an audio signal. In a latest paper [9], Dosiek reported the ENF extraction from digital recordings using Frequency Demodulation. In this approach, signal is modulated to zero frequency and analyze the signal.

One of the main characteristics of ENF signal is that it is sinusoidal in nature and it contains the odd harmonics of the fundamental frequency. Empirical Wavelet Transform (EWT) [10] is an AM-FM signal decomposition technique similar to Empirical Mode Decomposition (EMD) [13]. EWT algorithm decomposes the signal into number of modes by fragmenting the spectrum using matched band pass filtering concept. EWT captures the frequency boundaries according to the specified number of modes. Explicitly fixing the modes helps in capturing the harmonics in the signal effectively. However, one disadvantage of this approach is that, EWT cannot pick the weak harmonics or low-power signals. This paper resolved this issue by fixing the frequency boundary of desired modes of AM-FM signal. Boundary fixing helps to apprehend the mode analogous to ENF. Hilbert transform [11] was applied on extracted mode to obtain the instantaneous frequency and amplitude. The results showed that the instantaneous frequency obtained is equivalent to ENF (50Hz). Input signal was reconstructed by summing up all the bands together. This paper proposes a EWT approach for the extraction of 50Hz frequency component from an audio signal and this extracted signal can be matched against a reference signal extracted from PMU signal at home PDC to ensure the integrity of the signal. Hilbert transform is used to obtain the instantaneous frequency and amplitude of the extracted signal.

2. Electrical Network Frequency

Electrical Network Frequency is a voltage frequency captured along with a signal when it is recorded with a device connected to the power system. This frequency is unique across the network. The signal may also capture the ENF when the recording is exposed to an electromagnetic field or other power supplies [12]. Pattern of ENF is same in all the networks and hence we can say that ENF is stable and unique across the grid. ENF is also useful in identifying the time and location of the recorded data. ENF criterion can be used to evaluate the authenticity of a power system signal by comparing the extracted ENF with the previously stored ENF in the database. Instantaneous voltage from a power system can be written mathematically as,

$$v(t) = A\cos\left(2\pi f_C t + \theta(t)\right) \tag{1}$$

Where $v(t)$ represents the instantaneous amplitude, A is the amplitude of the signal, f_c represents the fundamental frequency and $\theta(t)$ is the phase of the signal at time t. Theoretically ENF is constant across the network. However, a slight change occurs due to some interference. A real time ENF can be expressed as [1],

$$f = (50 + \Delta f)\,Hz \tag{2}$$

Where Δf is a small change in frequency.

3. Empirical Wavelet Transform

Jerome Gills proposed a method [10] to extract the IMF using a wavelet filter bank. These Matched Band pass filters are built around the peaks in the frequency spectrum. The signal is segmented into N frames in frequency domain depending on the peaks. Perfect reconstruction of the signal is possible with these filters around each peak and IMFs are obtained by taking the inverse wavelet transform. In this method, a frame is marked with boundaries ω_{n-1}, ω_n and can be denoted as $\Lambda_n = [\omega_{n-1}, \omega_n]$ and each $\omega_n \in [0, \pi]$ and each partition is having a width of τ_n. Littlewood-Paley and Mayer's wavelets are used for constructing filter banks.
The scaling and empirical wavelet functions are defined as,

$$\hat{\phi}(\omega) = \begin{cases} 1 & \text{if } |\omega| \le \omega_n - \tau_n \\ \cos[\frac{\pi}{2}\beta(|\omega| - \omega_n + \tau_n)] & \text{if } \omega_n - \tau_n \le \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$\hat{\psi}_n(\omega) = \begin{cases} 1 & \text{if } \omega_n + \tau_n \le |\omega| \le \omega_{n+1} - \tau_{n+1} \\ \cos[\frac{\pi}{2}\beta(\frac{1}{2\tau_n}(|\omega|) - \omega_{n+1} + \tau_{n+1})] & \text{if } \omega_{n+1} - \tau_{n+1} \le |\omega| \le \omega_{n+1} + \tau_{n+1} \\ \sin[\frac{\pi}{2}\beta(\frac{1}{2\tau_n}(|\omega|) - \omega_n + \tau_n))] & \text{if } \omega_n - \tau_n \le |\omega| \le \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The cosine function finds the right of the mode and sine function finds the left of the mode. Therefore, this filter is able to perform the exact reconstruction. Sine and cosine functions are fitted by an arbitrary function,

$$\beta(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1 & \text{if } x \ge 1 \end{cases} \quad \text{and } \beta(x) + \beta(1-x) = 1, \ \forall x \in [0,1] \tag{5}$$

By the definition of EWT, for a function f the details coefficients are obtained by taking the inverse of convolution between f and $\hat{\psi}_n$.

$$\begin{aligned} W_f^{\varepsilon}(n,t) &= \langle f, \psi_n \rangle \\ &= \int f(\tau)\overline{\psi_n(\tau - 1)d\tau} \\ &= (\hat{f}(\omega)\overline{\hat{\psi}_n(\omega)})^{-1} \end{aligned} \tag{6}$$

The approximate coefficients are obtained by taking the inverse of convolution operation between f and ϕ_n.

$$\begin{aligned} W_f^{\varepsilon}(0,t) &= \langle f, \phi_1 \rangle \\ &= \int f(\tau)\overline{\phi_1(\tau - 1)d\tau} \\ &= (\hat{f}(\omega)\overline{\hat{\phi}_1(\omega)})^{-1} \end{aligned} \tag{7}$$

Therefore the signal $f(t)$ can be reconstructed as

$$f(t) = W_f^{\varepsilon}(0,t) * \phi_1(t) + \sum_{n=1}^{N} W_f^{\varepsilon}(n,t) * \psi_n(t) \\ = (\hat{W}_f^{\varepsilon}(0,\omega)\hat{\phi}_1(\omega) + \sum_{n=1}^{N} \hat{W}_f^{\varepsilon}(n,\omega)\hat{\psi}_n(\omega))^{-1} \quad (8)$$

The empirical mode f_k is given as

$$f_0(t) = W_f^{\varepsilon}(0,t) * \phi_1(t), \\ f_k(t) = W_f^{\varepsilon}(k,t) * \psi_k(t) \quad (9)$$

Fig. 1 represents the Boundaries obtained after portioning the frequency spectrum with Littlewood-Paley and Mayer's wavelets.

4. Results and Discussion

An audio signal embedded with a 50Hz signal was analyzed and is shown in Fig. 2.

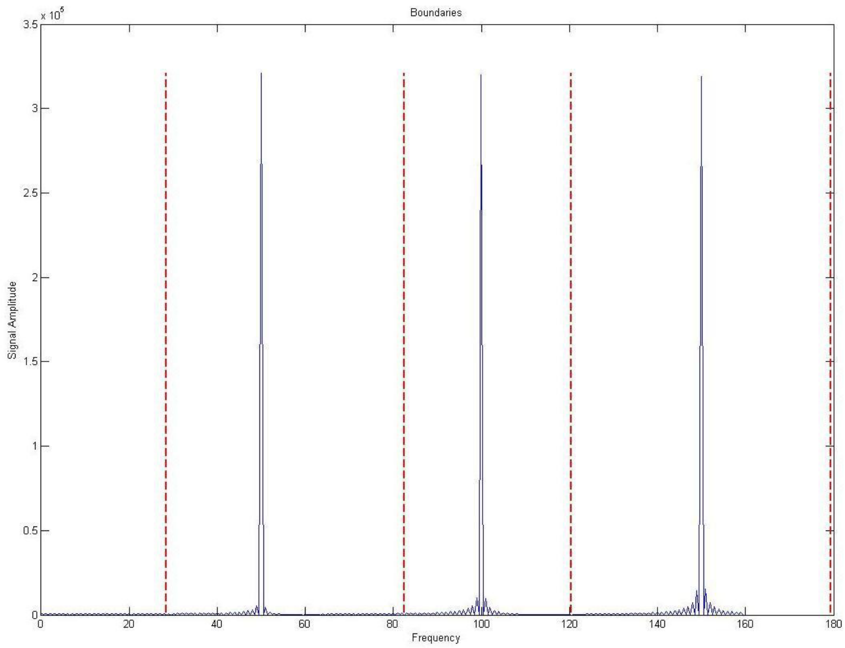

Fig. 1. Boundaries obtained after partitioning the frequency spectrum with Littlewood-Paley and Mayer's wavelets.

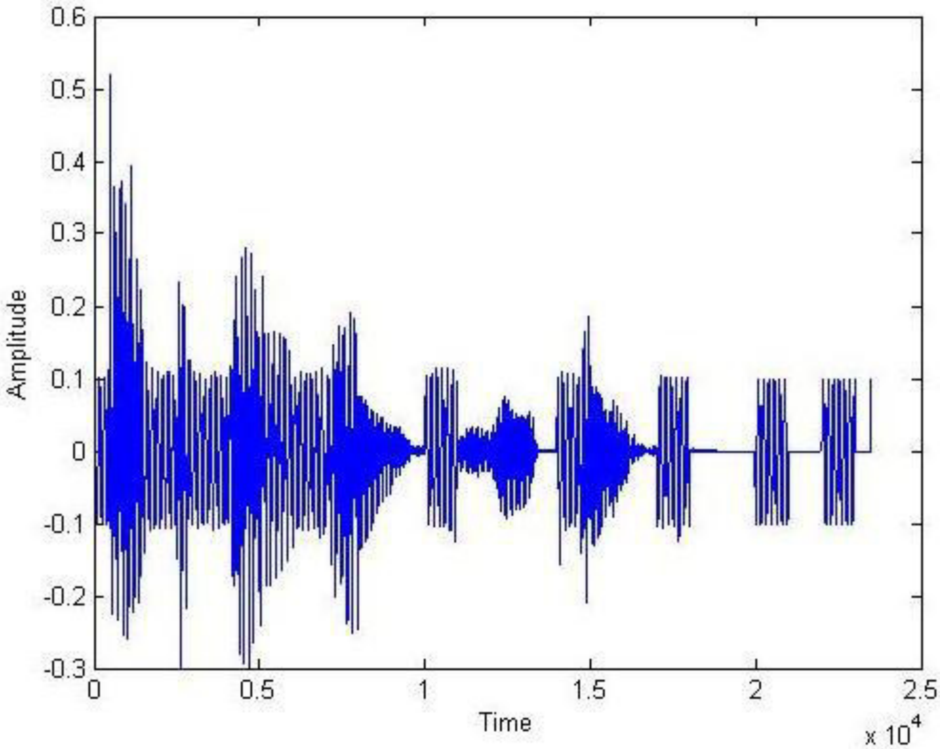

Fig. 2. Input audio signal embedded with 50Hz frequency

Empirical wavelet transform was applied on this input signal *f* and the number of modes to be extracted was fixed to 3. Number of boundaries are fixed around the spectrum peaks in the frequency spectrum. This can be easily identify by checking the frequency spectrum. Frequency domain boundaries are computed by extracting the local minima between two local maxima. A filter bank was implemented with Meyer's wavelet [14] for the obtained frequency boundaries. Fig. 3 shows the modes obtained after applying EWT on the input signal. It can be inferred that the first mode corresponds to a dc component as its amplitude value is approximately equal to zero (i.e., in the order of 10^{-6}). Fig. 4 gives an idea about the detected frequency domain boundaries. Two boundaries are obtained around 50Hz frequency. So, it is easy to segment out that particular mode. Instantaneous magnitude and frequency of this mode was obtained by applying Hilbert transform. Hilbert transform of a signal is computed as [15],

$$\hat{x}(t) = H\left[x(t)\right] = \frac{p.v}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{1-\tau}d\tau \tag{10}$$

Where $p.v$ is the principal value.

Fig. 5 shows the instantaneous amplitude and frequency of the extracted 50Hz signal. Instantaneous frequency is equivalent to 50Hz with a little oscillation. This extracted ENF signal can be cross correlated with a reference PMU data to assess the authenticity of the audio signal. Tampering an audio signal may change the characteristics of the ENF signal. Therefore, integrity or authenticity of an audio signal can be ensured by pair-wise cross-correlation between the extracted ENF and a reference ENF. Cross-correlation is the process of comparing two independent signals. Pair-wise Normalized cross-correlation between two signals is given as [16],

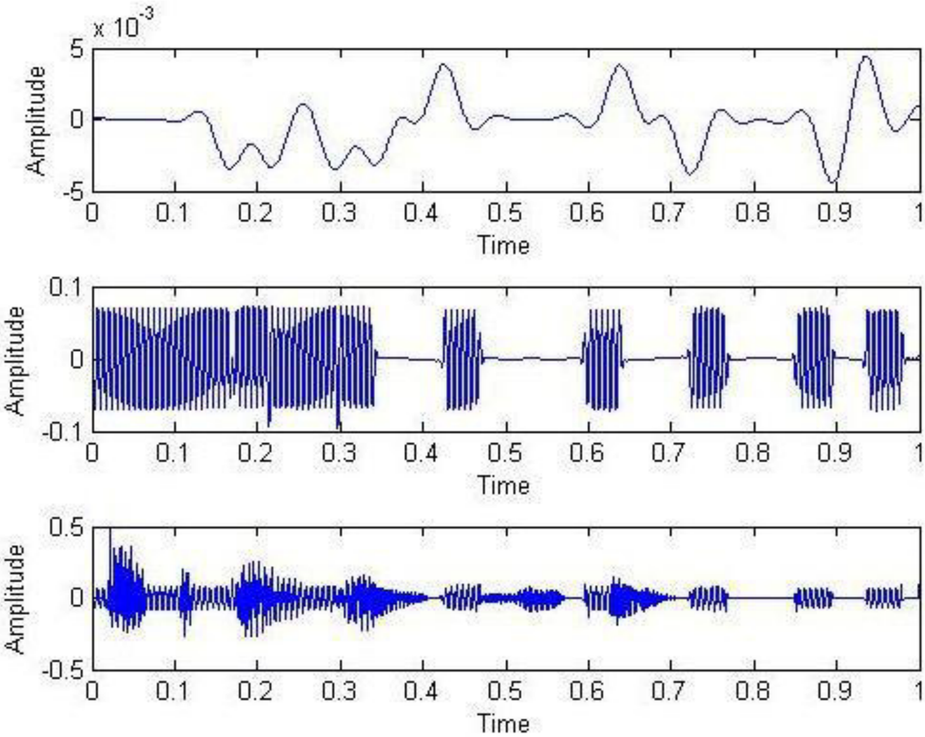

Fig. 3.Modes obtained after applying EWT on input signal

$$\rho = \frac{\sum_{i=0}^{N-1}\left(x(i)-\mu_x\right)\left(y(i)-\mu_y\right)}{\sqrt{\sum_{i=0}^{N-1}\left(x(i)-\mu_x\right)^2 \sum_{i=0}^{N-1}\left(y(i)-\mu_y\right)^2}} \qquad (11)$$

Where $x(i)$ and $y(i)$ are the instantaneous frequencies of the extracted ENF from audio signal and the reference ENF measured by PMU respectively. μ_x and μ_y are the mean values of $x(i)$ and $y(i)$ respectively. N is the length of the digital recording.

PMU data obtained from Texas Synchrophasor Network is used as a reference PMU data. ENF is fixed to 60Hz in USA and hence 60Hz is taken as reference ENF. For the experimental purpose, a 60Hz frequency is embedded in the audio signal that is created based on reference PMU data. 60Hz component was extracted efficiently using EWT and its instantaneous frequency was estimated using Hilbert transform. Cross-correlation between this instantaneous frequency and ENF obtained from the PMU is computed. Cross-correlation value is obtained as 1 which shows that both signals are highly correlated. When cross-correlation is above a threshold (say $\tau > 0.8$), the extracted ENF is matching with reference data and hence the inference is that ENF signal is not tampered. This procedure is really helpful in determine the authenticity of an audio signal. This correlation can be done only by PMU data and audio signal taken at the same time. This correlation procedure can be used for ensuring the authenticity an audio signal.

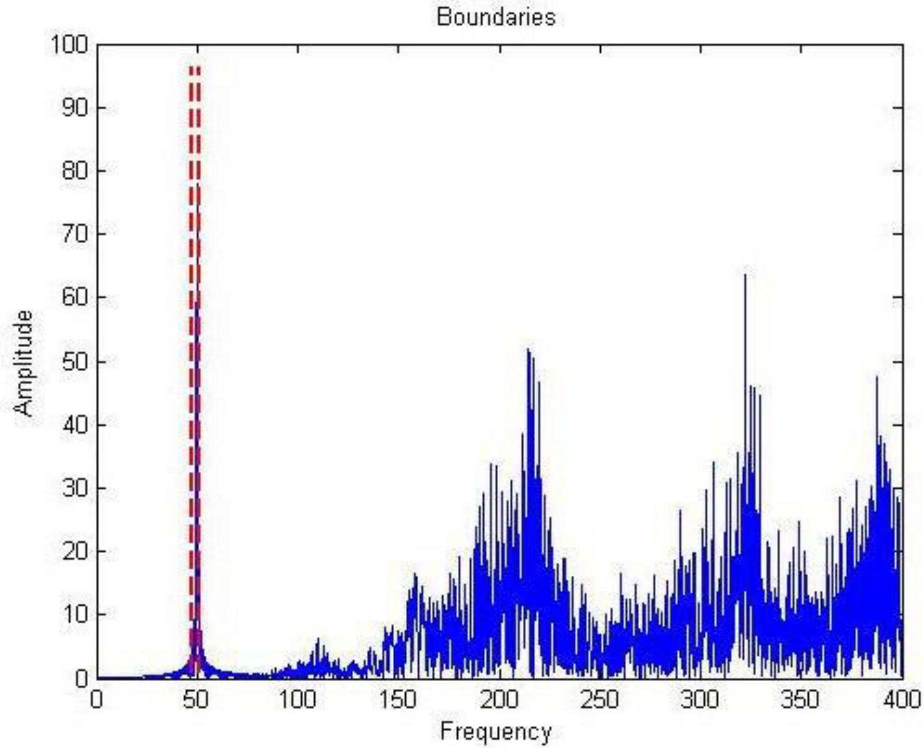

Fig. 4. Frequency domain boundaries for the input audio signal

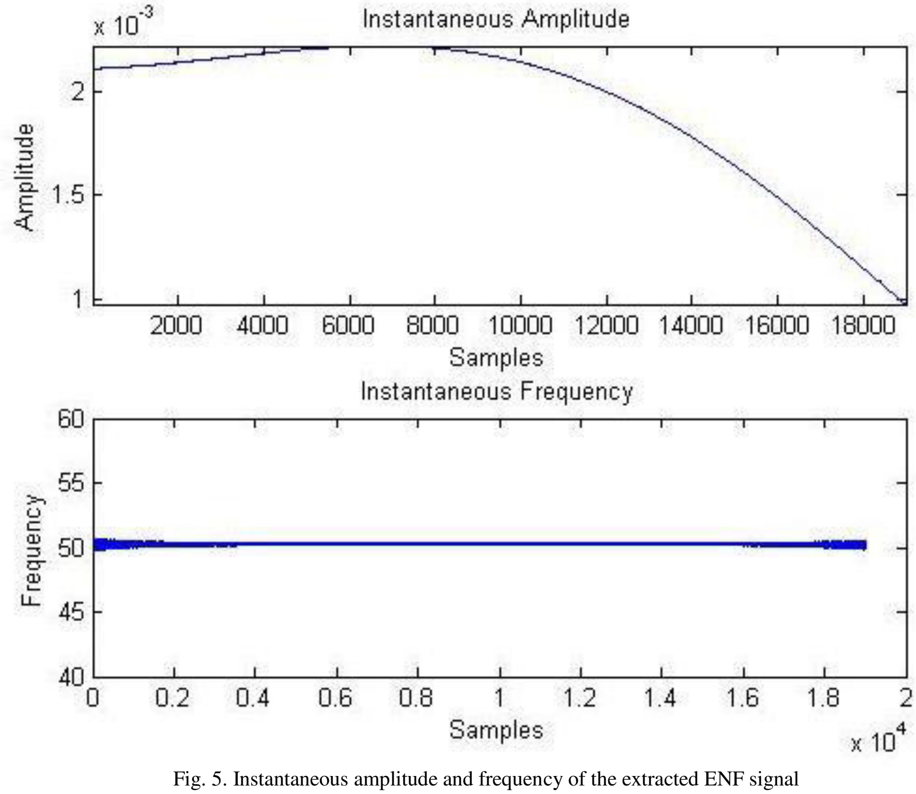

Fig. 5. Instantaneous amplitude and frequency of the extracted ENF signal

5. Conclusion

ENF extraction from an audio signal is considered as a powerful tool in digital forensics. Estimating the embedded ENF from an audio signal is efficiently done using Empirical Wavelet Transform. Hilbert transform was applied on the extracted ENF to compute its instantaneous frequency and amplitude. ENF measured from Texas Synchrophasor Network PMU data is then cross-correlated with the extracted ENF from the audio signal. A cross-correlation value of 1 indicates that both signals are highly correlated. This procedure can be effectively utilized in the authentication of audio signals. Every digital recording is embedded with ENF which can be correlated with the power system ENF to check whether any tampering is done on the record. Cross-correlation value of 1 signifies the high correlation. If cross-correlation between two signals is above a threshold (say 0.8), both signals are equivalent. This paper also resolved the disadvantage of EWT to pick weak harmonics by boundary fixing in frequency domain.

References

[1] Grigoras, Catalin. "Digital audio recording analysis–the electric network frequency criterion." International Journal of Speech Language and the Law12.1 (2005): 63-76.

[2] Grigoras, Catalin. "Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis." Forensic Science International167.2 (2007): 136-145.

[3] Kajstura, Mateusz, Agata Trawinska, and Jacek Hebenstreit. "Application of the Electrical Network Frequency (ENF) Criterion: A case of a digital recording."Forensic science international 155.2 (2005): 165-171.

[4] Varodayan, David, Yao-Chung Lin, and Bernd Girod. "Audio authentication based on distributed source coding." Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. IEEE, 2008.

[5] Farid, Hany. "Detecting digital forgeries using bispectral analysis." (1999).

[6] Yang, Rui, Zhenhua Qu, and Jiwu Huang. "Detecting digital audio forgeries by checking frame offsets." Proceedings of the 10th ACM workshop on Multimedia and security. ACM, 2008.

[7] Liu, Yuming, et al. "Wide-area frequency as a criterion for digital audio recording authentication." Power and Energy Society General Meeting, 2011 IEEE, 2011.

[8] Ojowu, Ode, et al. "ENF extraction from digital recordings using adaptive techniques and frequency tracking." Information Forensics and Security, IEEE Transactions on 7.4 (2012): 1330-1338.

[9] Dosiek, Luke. "Extracting Electrical Network Frequency From Digital Recordings Using Frequency Demodulation." Signal Processing Letters, IEEE22.6 (2015): 691-695.

[10] Gilles, Jerome. "Empirical wavelet transform." Signal Processing, IEEE Transactions on 61.16 (2013): 3999-4010.

[11] Poularikas, Alexander D., ed. Transforms and applications handbook. CRC press, 2010.

[12] Bykhovsky, Dima, and Asaf Cohen. "Electrical network frequency (ENF) maximum-likelihood estimation via a multitone harmonic model." Information Forensics and Security, IEEE Transactions on 8.5 (2013): 744-753.

[13] Huang, Norden E., et al. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 454. No. 1971. The Royal Society, 1998.

[14] Vetterli, M., and TA Tony Verma. "Filterbank implementation of meyer's wavelet." EE392G Stanford University (1998).

[15] Huang, Norden E., et al. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 454. No. 1971. The Royal Society, 1998.

[16] Garg, Ravi, Avinash L. Varna, and Min Wu. "Seeing ENF: natural time stamp for digital video via optical sensing and signal processing." Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011.

[17] Grigoras, Catalin. "Applications of ENF analysis in forensic authentication of digital audio and video recordings." Journal of the Audio Engineering Society57.9 (2009): 643-661.