Feasibility of using Electrical Network Frequency fluctuations to perform forensic digital audio authentication

By

Tarek El Gemayel

A thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements for the

M.A.Sc degree in Electrical and Computer Engineering

School of Electrical Engineering and Computer Science

University of Ottawa

Statement of Contribution of Collaborators

I am the sole author of all the sections of this thesis. My supervisor, Dr. Martin Bouchard of the School of Electrical Engineering and Computer Science, University of Ottawa, supervised my work throughout the M.A.Sc. program and provided editorial corrections.

Abstract

Extracting the Electric Network Frequency (ENF) fluctuations from an audio recording and comparing it to a reference database is a new technology intended to perform forensic digital audio authentication. The objective of this thesis is to implement and design a range of programs and algorithms for capturing and extracting ENF signals. The developed C-program combined with a probe can be used to build the reference database. Our implementation of the Short-Time Fourier Transform method is intended for the ENF extraction of longer signals while our novel proposed use of the Autoregressive parametric method and our implementation of the zero-crossing approach tackle the case of shorter recordings. A Graphical User Interface (GUI) was developed to facilitate the process of extracting the ENF fluctuations. The whole process is tested and evaluated for various scenarios ranging from long to short recordings.

Sommaire

Extraire les fluctuations des fréquences du réseau électrique (FRE) à partir d'un enregistrement audio et les comparer à une base de données de référence, est une nouvelle technologie permettant l'authentification légale de fichiers audio. L'objectif de cette thèse est d'implémenter et de concevoir une gamme de programmes et d'algorithmes pour l'extraction de ces signaux FRE. Combiné avec une sonde, le programme-C développé permet de construire une base de données de référence. Notre mise en œuvre de la méthode de la Transformée de Fourier à Court Terme est destinée à l'extraction des FRE d'un signal à longue durée, tandis que l'utilisation nouvelle que nous proposons de la méthode Autorégressive paramétrique et notre mise en œuvre de l'approche des croisements par zéro sont utiles pour les fichiers audio de courte durée. Une Interface Graphique (IG) est développée et l'ensemble du traitement est testé et évalué pour divers scénarios d'enregistrements, longs et courts.

Acknowledgements

I would like to express the deepest appreciation and gratitude to my supervisor, Dr. Martin Bouchard, for his continuous support and his immense motivation skills. It was through his unforgettable and wonderful undergraduate classes that I was introduced to the world of Digital Signal Processing. He continually and influentially provided me with the spirit of adventure throughout my Master's program, even though at times, it seemed more like a dead end.

I would also like to thank Dr. Matthew Case, from the Royal Canadian Mounted Police, for giving me the chance to work on this project. His continuous involvement in our team meetings was of great assistance. His cheerful enthusiasm and ever-friendly attitude always gave me hope to never give up.

Finally, I feel privileged to have a family, parents and a brother, whose love and support has kept me going through difficult times and friends who stood by my side and supported me throughout my endeavors.

Table of Contents

List of Tables

List of Figures

List of Symbols

B = Magnetic flux density

d = Distance between two wires

H = Magnetic field

$H(z)$ = Filter frequency response

$h(n)$ = Windowed filter impulse response

$h_d(n)$ = Ideal low-pass filter impulse response

$h_{\text{mod}}(n)$ = Modulated filter impulse response to bias frequency

I = Current

μ_0 = Permeability of free space

μ_v = Mean of the noise

ρ_{xy} = Cross-correlation coefficient between signal x and y

R_x = Correlation matrix

$r_x(k)$ = Correlation coefficient

σ_x = Root mean squared value of signal x

$S_{xx}(e^{j\omega})$ = The useful signal's power spectral density

$S_{vv}(e^{j\omega})$ = The noise signal's power spectral density

τ_i = Time index of zero-crossing

$V(z)$ = Noise frequency response

$w(n)$ = Blackman window function

ω_c = Cut-off frequency

w_k^* = Auto-regressive coefficients

$X(z)$ = Useful signal frequency response

List of Acronyms

AC = Alternating Current

ACE = Automatic Control Error

AES = Audio Engineering Society

AGC = Automatic Generator Control

AR = Autoregressive

AVAU = Audio and Video Analysis Unit

ENF = Electrical Network Frequency

FFT = Fast Fourier Transform

FIR = Finite Impulse Response

GB = Gigabyte

GUI = Graphical User Interface

Hz = Hertz

IDE = Integrated Development Environment

IIR = Infinite Impulse Response

MB = Megabyte

MSE = Minimum Squared Error

NCMF = National Center for Media Forensics

NERC = North American Electric Reliability Corporation

PC = Personal Computer

PCM = Pulse Code Modulation

PSD = Power Spectrum Density

RCMP = Royal Canadian Mounted Police

STFT = Short-Time Fourier Transform

TB = Terabyte

UPS = Uninterruptable Power Supply

USA = United States of America

V = Volts

1. Introduction

1.1. General introduction and background

Nowadays thanks to the impactful technological advancements the world has faced over the past few years, digital audio and video recordings are being used as evidence in legal proceedings, thus bringing the subject of authenticity into question. The use of analog recording devices such as magnetic tapes and vinyl discs has been on the verge of extinction and is mainly overshadowed by the digital world. Indeed, CDs, USBs, hard drives, digital recorders, cellphones and many more are now utilized to capture and store recordings. The ease of use of computers and the abundant editing software solutions, with the help of tutorials, forums and detailed manuals, have given the capability to almost anyone to edit and modify digital audio recordings. If handled carefully, digital files can be meticulously altered and manipulated in order to be undetectable to the eye or the ear, whereas modified analog recordings leave distinctive features along the tape (e.g. blanks, cuts) which can easily be seen or heard, hence labeling the audio recording as unauthentic [2].

According to the Oxford English Dictionary, labeling something as “authentic” means that it is “real, actual, ‘genuine’ ” and “really proceeding from its reputed source or author; of undisputed origin” [3]. In addition, the Audio Engineering Society (AES) defines an authentic recording as being “a recording made simultaneously with the acoustic events it purports to have recorded, and in a manner fully and completely consistent with the method of recording claimed by the party who produced the recording; a recording free from unexplained artifacts, alterations, additions, deletions, or edits” [4]. Therefore, performing forensic digital audio authentication can be ultimately

perceived as the act of identifying the genuineness of the audio while applying scientific know-how to solve legal issues. The term “forensic” derives from the word “forum” which first appeared in Latin meaning “the public place or market-place of a city” and “in ancient Rome the place of assembly for judicial and other public business” [5] which translates to speaking the truth in public or before a panel of judges. Modern forensic science dates back to the late 1800’s when Paul Kirk and Edmund Locard proposed that forensics is the study of comparison [6]. Since then the application of forensics has spread enormously and is now seen in various study fields such as pathology, anthropology, toxicology, odontology, arts, engineering and more precisely digital audio forensics [7]. The judiciary panel often calls upon audio experts in order to examine, analyze and “dissect” digital recordings in order to deliver a proper judgment to the court of law pertaining to a specific crime.

One has got many reasons for modifying a recording, whether it is for planning a murder, evading a murder, escaping monetary frauds or creating propaganda. For example, a police officer recorded a gang selling weapons where the latter was arrested but later claimed that the police had manually created the recording from various files. The gang members were incarcerated for a total of 33 years after a group of forensic experts were asked to verify the recording’s authenticity [8].

Although it is impossible to put a stop on digital forgery, the best approach to overcome such situation is to verify if the digital file has been tampered with. A couple of great articles addressing the subject of audio authenticity can be retrieved despite the fact that the matter has not been extensively tackled. Most notably, the article on analog recordings written by Bruce E. Koenig in 1990 [9] describes that “critical listening,

physical inspection, magnetic development, and signal processing with waveforms, narrow-band spectra, and spectrograms are the most common laboratory procedures in authenticity determinations" [9]. In most scenarios the forensic expert will require that all devices and components employed during the recording process be present for the investigation as well as the various recordings (authentic/unauthentic and original/copied recordings) presented in court. Magnetic recording and copying is based on the principle of converting electrical energy into magnetic energy and vice-versa, which is most commonly known as transduction. Therefore, it is necessary to perform the forensic analysis on the original analog file because creating copies would yield to a loss in quality in the subsequent files.

A second article from Koenig and Lacey published in 2009 [10] covers the situation of digital footages. The methodologies previously elaborated underwent minor modifications and the authors now propose a system based on 11 procedures (i.e. evidence marking, physical inspection, digital data imaging, playback/conversion optimization, critical listening, high-resolution waveform analysis, narrow-band spectrum analysis, spectrographic analysis, digital data analysis, miscellaneous techniques and work notes/reporting [10]) in order to authenticate a recording. Digital reproduction can produce exact replicas of the original file simply by copying the discrete binary information from one file to another. Distinguishing between the original/authentic and modified/unauthentic can sometimes be impossible, thus creating a problem for forensic analysts. For this reason a large number of other authentication examination methods have been developed, some of them undertaking the subject of Electrical Network Frequency (ENF) [1, 11-24].

1.2. Thesis objectives

The scope of this thesis consists of analyzing and studying the feasibility of using Electric Network Frequency (ENF) fluctuations in order to authenticate digital audio recordings. The authentication process begins by extracting and analyzing the electrical fluctuations from both an audio recording and the electrical network reference. The reference grid signal is captured through a probe plugged into a power outlet and is then sent to a personal computer (PC) with sound board which will record the signal via a self-developed C-program, thus constructing the grid database. The ENF extraction requires the implementation of a downsampling algorithm followed by the evaluation of the power of the 60Hz component. The latter can be achieved using numerous approaches, such as the Short Time Fourier Transform (STFT) method, the zero-crossing method and our proposed Auto-Regressive parametric method. A Graphical User Interface (GUI) program is designed using MATLAB and is intended to handle the case of long recordings (e.g., 1 hour in length) as well as shorter recordings (e.g., 5 minutes in length). The end results will provide the examiner with detailed evidence allowing him to decide whether the audio recording, in question, is falsified or genuine. To be more precise, the information gathered from the ENF signal will eventually be utilized to determine the date and location at which the recording has been made and the position at which the recording has been modified, in case any discrepancies are found. This approach is tested in the thesis for the case of altered and unaltered recordings, each having various lengths and captured with multiple devices (battery powered or mains powered), using the methods previously mentioned. Ultimately, the advantage of obtaining the best possible

results resides in selecting the appropriate and suitable method of extracting the ENF depending on the characteristics of every recording.

During the subsequent sections of the thesis, a background on ENF signals and how they operate along a transmission line is presented in Section 2. In addition, the ENF capturing process for the grid file and for the audio file is explained later in the same section. The software programs employed for the purpose of the thesis are presented in Section 3. Section 4 presents the three methods (i.e. STFT, AR and zero-crossing) that are used for the ENF extraction process from both the audio and grid files. The building of an ENF database from grid ENF files originating from the probe is explained in Section 5. Lastly the various tests and results applied to the methods described in Section 4 are explained, analyzed and presented in Section 6.

2. Electrical Network Frequency

This section presents the topic of ENF that is later implemented in order to authenticate a digital recording. It is also recommended to consult additional references regarding other methods intended to authenticate digital recording, such as the ones elaborated by Bruce E. Koenig in [10]. First developed by Catalin Grigoras, director of the National Center for Media Forensics (NCMF) and assistant professor at the University of Colorado, Denver, the ENF criterion is based on the concept that small disturbances occur over time in the power grid line and can be extracted from the audio recording, and then compared to a reference database.

2.1. How ENF works

Generators rotating at the speed of 60 cycles per second in North America and 50 cycles per second in Europe produce an Alternating Current (AC) that travels along the power grid transmission lines. Ideally the AC frequency is fixed to a stable and constant value of 60Hz or 50Hz, depending on the region. However electricity production is contingent on power demand. The North American Electric Reliability Corporation (NERC) states that: “unlike water or gas, electricity cannot be stored. It must be generated as it is needed, and supply must be kept in balance with demand” [25]. Differences in supply and demand will introduce small disturbances and variations in the AC causing its frequency to vary over time. Higher demand will cause the frequency to drop while lower demand (or higher production) yields to a higher frequency value. The fluctuations introduced in the grid are random, non-predictable and most importantly are uniform throughout the entire transmission line of each interconnection. The Canadian territory is covered by three major interconnections (whereas North America is

comprised of four major interconnections), which in their turn are also tied to the United States of America's (USA) interconnections (please refer to the map in [26]):

- Western interconnection covers British Colombia and Alberta as well as the western part of the USA;
- Eastern interconnection covers the eastern provinces extending to the Saskatchewan/Alberta borders (excluding Quebec) as well as the eastern part of the USA;
- Texas interconnection covers the state of Texas;
- Quebec interconnection is commonly included in the eastern interconnection, however it has its own transmission grid for physical reasons [27, 28].

Nine regional councils (including Alaska) are spread out across the NERC territory and they are "responsible for monitoring compliance of the registered entities within their regional boundaries, assuring mitigation of all violations of approved reliability standards and assessing penalties and sanctions for failure to comply" [29]. In addition, these councils must monitor the flow of electricity, via balancing authorities connected to generators, and perform the proper adjustment to keep the frequency around its proper value of 60Hz. All the generators are synchronized with each other, meaning that they must all rotate in unison (i.e. at the same speed). Whenever a balancing authority dispatches a signal, all generators within the interconnection must regulate accordingly. However, when a generator rotates at a speed of ±2Hz with respect to the rest of the generators, the former can generate a considerable amount of heat leading to self-destruction [30]. On the eastern grid, the upper and lower limits to the frequency, yielding to the intervention of the balancing authority, are 60.02Hz and 59.98Hz (i.e.

error margin of ±0.02Hz) [27]. The balancing authorities are all based on an Automatic Generator Control (AGC) that rectifies any occurrence of frequency deviation. Subdivided into four controllers, the primary control will solve the small variations that occur within the matter of seconds, hence originating from secondary customer's loads. The secondary control computes an Area Control Error (ACE), which indicates the variations occurring within minutes, due to load changes originating from primary customers, and will then stabilize the frequency to its base value of 60Hz. The tertiary control tackles the variations occurring within a time frame of 10 minutes to an hour due to a load change originating from subtransmission customers [25]. Lastly the time controller keeps track of the time error accumulated from the ENF signal and sends out a notice to the balancing authorities whenever the threshold of 10 seconds over a period of 10 hours, has been reached [27, 31]. For example if the frequency is running at 60.004Hz then the clock will have an advance of 2.4 seconds after 10 hours (see Equation 2.1).

$$\frac{60.004-60}{60} * 10hrs * 3600secs/hr = 2.4secs \quad (2.1)$$

2.2. The authentication approach using ENF criterion

The process of authenticating an audio recording is divided into two main parts. The first one is based on capturing the ENF signal from the power outlet in order to construct

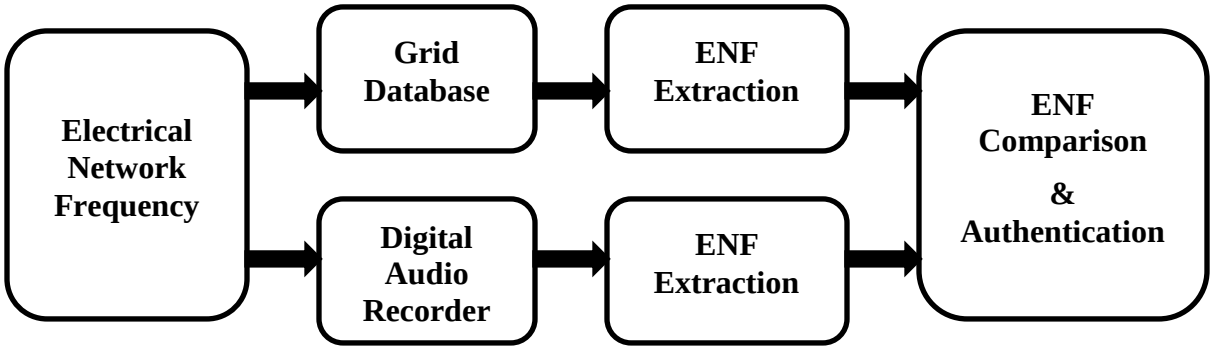

Figure 2.1 - Capturing the grid's ENF

the reference database, whereas the second part is based on capturing the ENF signal throughout the course of recording the audio file. Figure 2.1 presents the schematic of the general steps required in order to ultimately validate the recording.

2.3. Capturing the grid's ENF

The grid ENF will serve as the reference when it comes to confirming the recording's authenticity, or to identify the time and location of a recording. For this reason a few probes have been built, for the purpose of this thesis, based on the design in [1] and they can

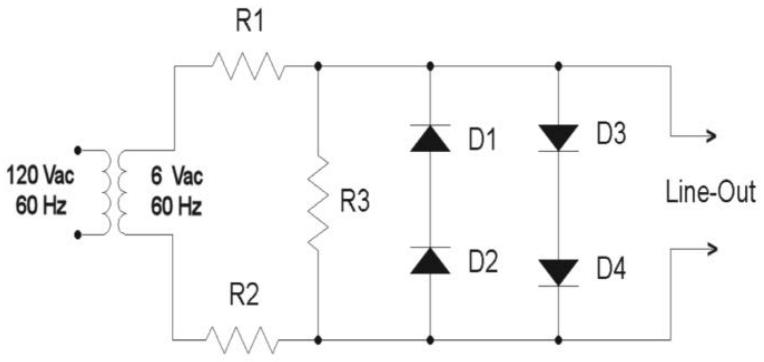

Figure 2.2 - Electrical circuit implemented for the design of the probe [1]

be used to convert the 120VAC electrical current into a ~1.2VAC peak to peak current which will serve as an input to a computer's microphone jack. The circuit shown in

Figure 2.2 is comprised of a 120VAC to 6VAC transformer followed by a voltage divider (R1=R2=15kΩ and R3=3170Ω) that lowers the voltage to around 1.2VAC peak to peak. The series of anti-parallel diodes (D1=D2=D3=D4=1N5246B) will protect the circuit from any random voltage spike above ~1.4V (i.e. ~0.7V cut-off voltage for each diode). A prototype of the probe is depicted in Figure 2.3. A RCA cable connects the probe's output to the computer's microphone jack input allowing the PC to record the incoming 60Hz ENF components, originating from the mains-powered outlet. The signal is recorded on the PC via a custom developed C-program utilizing 16bits per sample at 44.1kHz sampling rate. Section 3 of the thesis tackles the subject of the different software programs employed for this thesis.

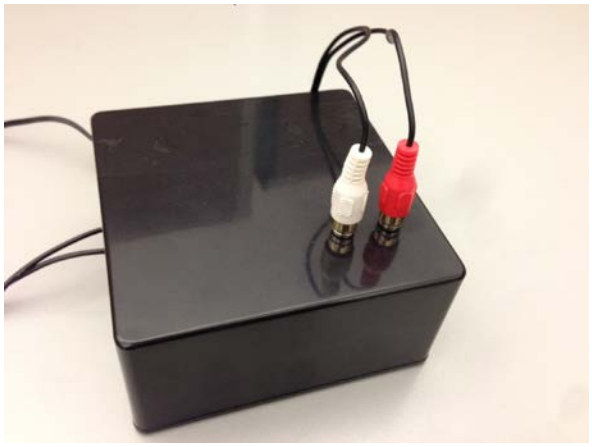

Figure 2.3 – Prototype of probe used for capturing the grid's ENF directly from the power outlet

It is necessary to mention that the plug needs to be connected to the electricity power socket (and not any other source of power) since the recording process must record any activity present on the electric network. In case there is a power outage the program will not detect any ENF signal and this can be quite useful during the testing phase. Due to the fact that Canada is divided into three electrical transmission networks, the RCMP is only required to constantly record the ENF signal on three separate computers.

2.4. ENF signals with digital audio recorders

Digital audio recorders are seen as devices that capture sound and then store the file in a discrete format. To be more precise, sound waves propagating in space are captured

by the device's microphone(s) (note: having multiple microphones on a device means that the concept of beam-forming can be applied, hence producing better sound quality and eliminating some unwanted background noise) and they are then sent to an anti-aliasing filter that will limit the signal's bandwidth, hence respecting the Nyquist theorem. Not band limiting the original signal could introduce distortions in the final digital signal due to the fact that the signal's original higher frequencies may be higher than half of the device's sampling rate. The filter's output is passed through an Analog-to-Digital Converter (ADC) whose objective is to discretize the analog signal considering the pre-determined sampling rate and bit-resolution. The sampling rate determines the number of samples required to model 1 second of audio and the bit-resolution determines the number of intervals present to model the signal's amplitude (i.e. having 16-bits resolution means that the amplitude is quantized to 65535 levels ranging from -32768 to 32767 or from 0 to 65535, see Equation 2.2 where N is the number of intervals and M the bit-resolution).

$$N = 2^M - 1 \tag{2.2}$$

The final result can be saved on a storage device and can easily be reproduced, as stated in Section 1.1. If a recording device is powered via the electricity outlet it is likely that some 60Hz ENF signal will be embedded in the recording [1, 17, 21, 32], for example as shown in Figure 2.4 for a mains-powered audio recording device. However it is also important to study the case where the device is powered through a battery source and yet will still include 60 Hz ENF fluctuations in the audio through electromagnetic waves propagation.

It is known that wirings and electrical components having conducting materials (such as copper) radiate magnetic fields in their vicinity. The magnitude of the magnetic field is proportional to the current's intensity and inversely proportional to the distance from the center of the wire.

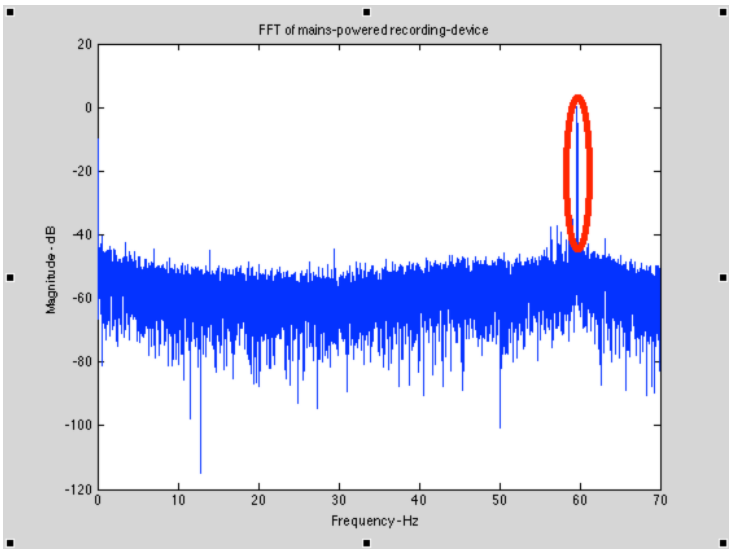

Figure 2.4 - Strong presence of 60Hz ENF component using mains-powered devices

Indeed a single and long wire carrying a current "*I*" produces a magnetic flux density "*B*", see Equation 2.4, or a magnetic field "*H*", see Equation 2.5, by applying the Biot-Savart Law described in Equation 2.3.

$$B = \frac{\mu_0 I}{4\pi} \int_{path} \frac{d\vec{l} \times \vec{r}}{r^3} \qquad 2.3$$

$$B(r) = \frac{\mu_0 I}{2\pi r} \qquad 2.4$$

$$B = \mu H \qquad 2.5$$

However in more complex situations such as house wiring, where the current powering the appliances flows in one separate wire than the current returning from that same appliance, the magnetic flux density at a certain distance is expressed using Equation 2.6, where "*d*" is the distances between both wires. In this case the magnetic

flux density decreases at a much faster rate than what is experienced having only one single wire.

$$B(r) = \frac{\mu_0 Id}{2\pi r^2} \qquad 2.6$$

In other scenarios where loops or coils are involved, most commonly found in appliances and transformers, the magnetic flux density is inversely proportional to the cube of the distance [33-36]. Since transmission lines, distribution lines and power stations are widespread throughout the world, it is often impossible for an electronic device to be fully isolated to the electromagnetic field that they generate. Therefore the recording devices powered by a battery source are also constantly susceptible to the exposure of electromagnetic waves (within a certain distance from nearby wires and appliances) and Figure 2.5 obtained from a battery-powered audio recording device shows that indeed a significant component is located at approximately around 60Hz.

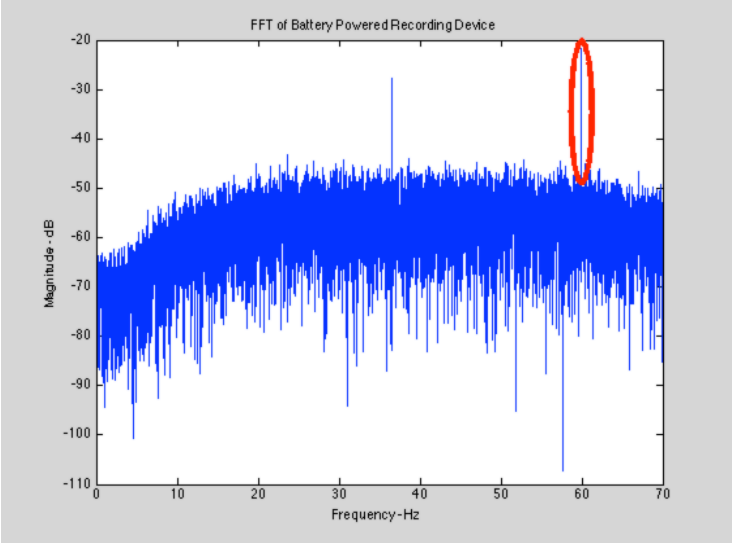

Figure 2.5 - Presence of 60Hz component using a battery-powered recording device

For the purpose of this thesis, the devices that were provided by the RCMP hence used for recording audio are the Olympus WS210-S (only battery-powered), Olympus DS-150 and Marantz recorder (can be battery-powered or mains-powered).

3. Simulation software

Two main commercial software environments have been utilized throughout the course of the thesis. The grid's capturing program was built on Microsoft Visual Studio, a C/C++ language programming software. However the program that detects, extracts and compares the ENFs from both signals was built using MATLAB.

3.1. Microsoft Visual Studio

Microsoft Visual Studio is an Integrated Development Environment (IDE) that allows a user to create a program in multiple languages. As stated in section 2.3, the probe's output is connected to a PC soundboard via a RCA cable. The developed program captures the incoming signal at a rate of 44.1kHz and will record it with 16 bits per sample resolution. That sampling rate could eventually be reduced to limit the amount of storage space required, all depending on the examiner's preference (this step requires a proper PC soundboard that can handle lower sampling rates).

The program will continuously record the incoming data and will save every 12 hours of information in a separate ".raw" binary file whose name is pre-defined according to the time at which the file is created. In addition, the C-program records the time-stamp of every recorded frame of signal in a separate ".raw" file, providing to the user the ability to determine the exact time and date of each recorded reference ENF frame (i.e. this will allow to verify if any frames were skipped during the recording process). The current frame size for the ENF reference recording is set to 10 seconds in the C-program. Using such a large value makes the system more robust to fluctuations in the recording caused

by the computer's operating system scheduling. Navigating through the ENF reference files is quite simple since their names have the following format, see Table 3.1.

	File name
Information recorded	recorded_year_day_month_time.raw *(e.g. recorded_2013_01_Mar_12'42"14_PM.raw"*
Time-stamp	time_year_day_month_time.raw *(e.g. time_2013_01_Mar_12'42"14_PM.raw"*

Table 3.1 – Name format used when recording the ENF and time-stamping each file

Furthermore a command prompt appears on the screen once the program is executed and will provide live visual information to the user about the recording process (i.e. time at which the grid ENF file and time-stamp file have been created). An example of the command prompt is shown in Figure 3.1. The capturing and recording process are terminated when the user closes the command prompt window.

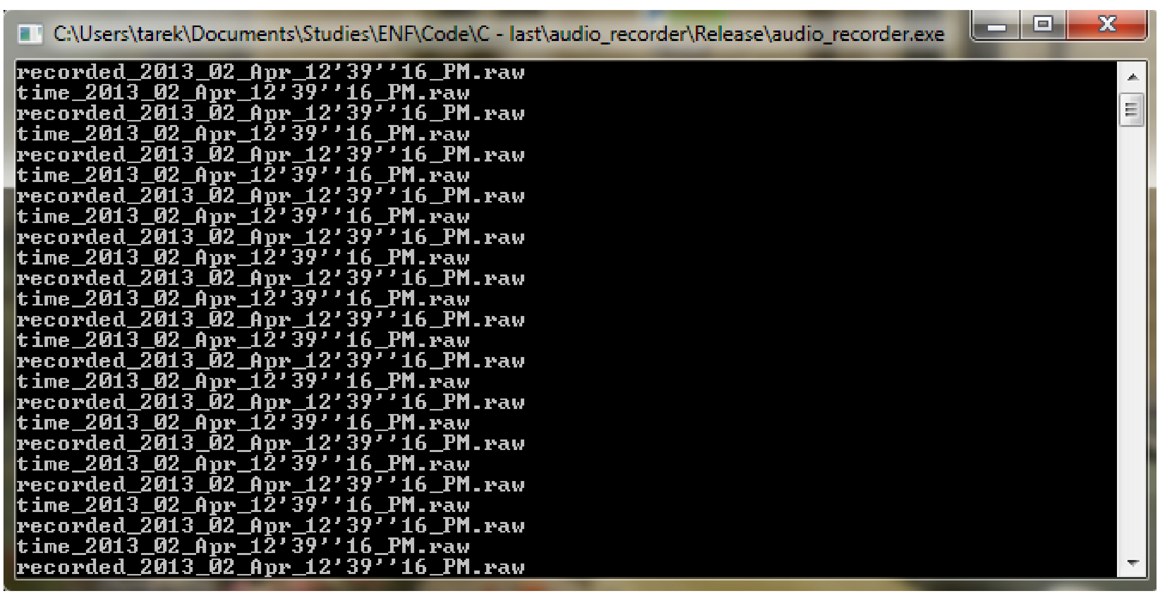

Figure 3.1 – Command prompt window indicating that the recording process is active and displaying the files' names created while using the format shown in Table 3.1

3.2. MATLAB by MathWorks

The ENF extraction and comparison process is performed on a self-developed GUI program built using the MATLAB software. The GUI provides the examiner with a wide range of options to determine the recording's authenticity. First the grid's ".raw" file must be loaded onto the program followed by the selection of an appropriate method to extract the 60Hz ENF signal (e.g., STFT or AR methods). If no option is selected and the examiner decides to extract the ENF, an error message pops-up demanding the selection of an extraction method, see Figure 3.2 for an example. Section 4 provides a detailed explanation of the various methods. The possibility of saving the ENF signal is available after computing the required calculations.

Figure 3.2 – Error message obtained if no method is chosen for extracting ENF from grid

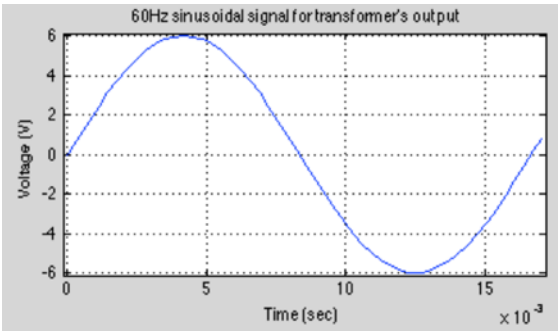

Figure 3.3 – Transformer's 60Hz output

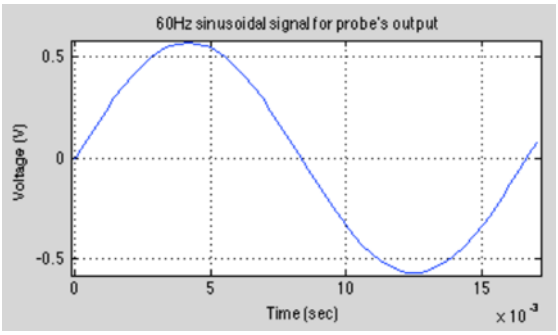

Figure 3.4 – Probe's 60Hz output

As for the audio recordings, similarly the user is required to select the proper parameters (i.e. sampling frequency of audio, method for extracting ENF and desired frequency for extracting ENF) for the corresponding ".wav" 16 bits Pulse-Code Modulation (PCM) coded audio file. Choosing MATLAB as the main software for authentication is an excellent choice. The software is equipped with integrated functions

as well as a strong mathematical base that gives the ability to easily manipulate matrices and plot figures.

Simulink, a design and modeling program of MATLAB, also helped to model the probe's electrical circuit, refer to Figure 2.2, and perform tests while obtaining great visual results of the simulated outputs for the 60Hz sinusoidal signal, see Figures 3.3, 3.4 and 3.5.

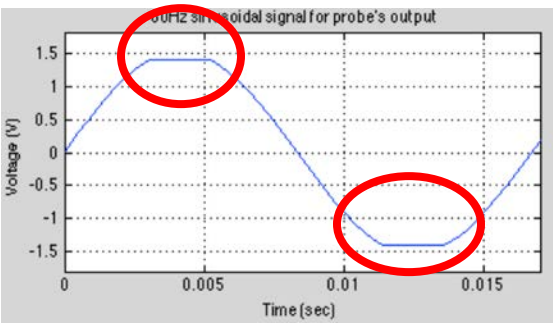

Figure 3.5 – Probe's output when voltage exceeds exceeds the diodes' limit of ~1.4V

This allowed verifying that the anti-parallel diodes protect the computer's soundcard from any power surge or voltage spike that occur on the network by clipping the signal's waveform whose voltage is above ~1.4V, see Figure 3.5.

A picture of the GUI that has been developed is shown in Figure 3.6.

3.3. Adobe Audition

Adobe Audition is an audio multimedia software developed by Adobe and is utilized throughout this thesis for verification purposes. The program is capable of plotting, in real time, the frequency graph of an audio file, see Section 6.4, providing simple and fast results. Adobe Audition is also used to alter the audio files that are employed in the thesis.

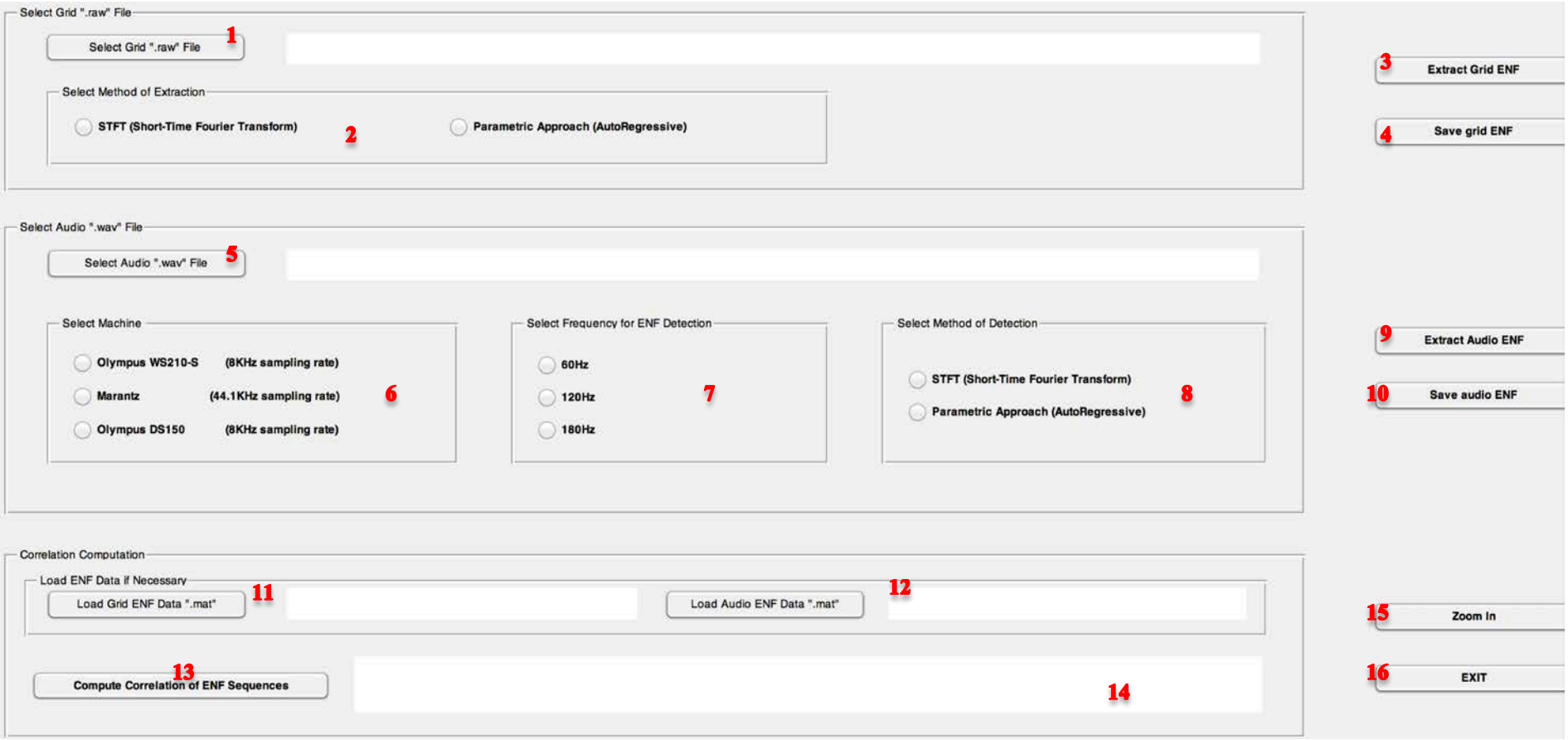

Figure 3.6 – GUI developed using MATLAB

1) Button to select grid file
2) Methods to extract grid ENF
3) Button to extract grid ENF
4) Button to save grid ENF
5) Button to select audio file
6) Device used to record audio
7) Frequency for ENF extraction
8) Methods to extract audio ENF
9) Button to extract audio ENF
10) Button to save audio ENF
11) Button to load grid ENF
12) Button to load audio ENF
13) Button to compute correlation
14) Result of correlation
15) "Zoom in" at match location
16) Exit button

4. ENF extraction

The process of authenticating a digital audio recording begins by the extraction of the ENF fluctuations from both the audio recording itself and the electrical network reference captured via the probe. According to previous works by Catalin Grigoras et al. [1, 17, 19-21], frequency domain Short-Time Fourier Transform or Fast Fourier Transform (FFT) methods can be used to evaluate the long term fluctuations of the ENF component (thus useful for longer audio signals). Conversely, the zero-crossing approach or our proposed AR modeling method can be used to evaluate the short-term fluctuations of the ENF components (thus potentially useful for shorter audio signals).

4.1. Short-Time Fourier Transform method

The STFT method and the FFT methods can actually be seen as equivalent, they only differ in the presentation of the results (i.e. Spectrogram for STFT and spectrum magnitude for the FFT). Both of them are based on a discrete Fourier Transform and are computed with FFTs in practice. More precisely, the STFT is defined as being the

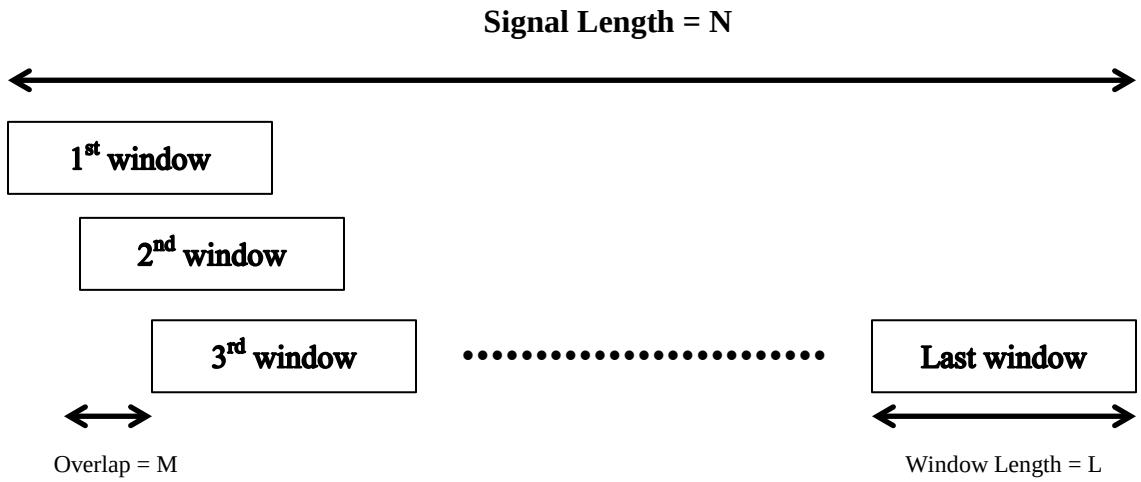

Figure 4.1 – Concept of the Short-Time Fourier Transform by applying a sliding window along the signal

magnitude Fourier Transform taken on a portion of the signal, by applying a sliding window of 200 seconds in order to cover the whole function, see Figure 4.1 for a visual representation.

The number of windows required to scan the entire signal is contingent on the signal length *N*, the window size *L* and the shift *M* applied between each consecutive window, see Equation 4.1.

$$\#windows = \frac{N - L + M}{M} \qquad 4.1$$

In addition, the result obtained can be seen as being a two-dimensional (i.e. frequency and time) representation of the signal if a spectrogram representation is used, or simply a sequence of magnitude spectra if the FFT magnitude representation is used instead.

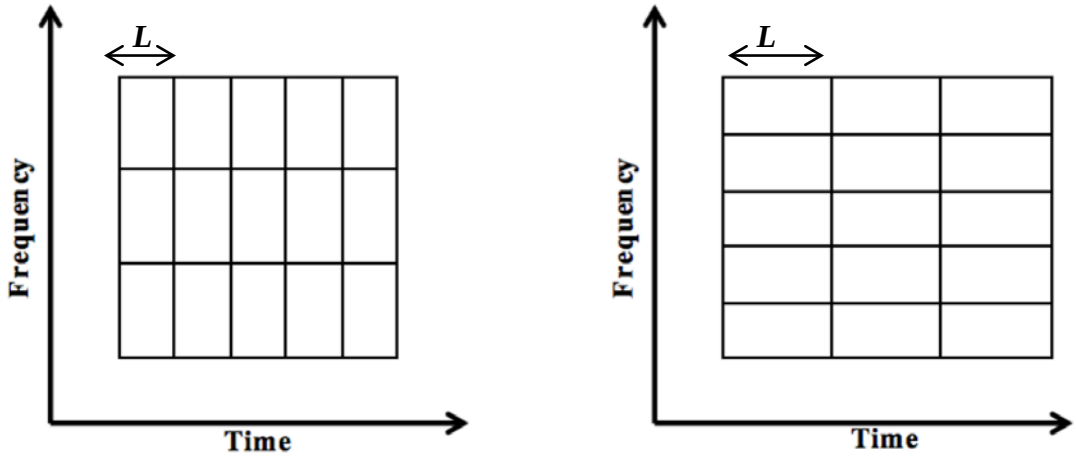

Figure 4.2 – Better time resolution than frequency resolution for shorter windows (left) and better frequency resolution than time resolution for longer windows (right)

Manipulating the window's length will yield to different resolutions in time and in frequency as seen in Figure 4.2. Longer windows will produce better frequency resolution due to the increase in the number of sample data and shorter windows will produce better time resolution. However, it can be quite strenuous deciding over which

value fits best the requirements. Clearly it would be ideal to obtain the best frequency and time resolution as possible with the same window, but this cannot be achieved in practice.

4.1.1. Digital audio recording file

In order to analyze the audio recordings, the user must load the corresponding 16 bits PCM ".wav" file obtained from a recording device using the "Select Audio '.wav' File" button on the GUI (see Figure 3.6), followed by the selection of the device used to record the audio. Both of the Olympus devices have a sampling rate 8kHz whereas the Marantz has a much higher sampling rate of 44.1kHz. Selecting the frequency for the ENF extraction is mandatory since it will determine at which frequency the ENF extraction will take place. In some cases the 60Hz ENF fluctuations might have been erased by the offender to erase the evidence or might have been covered by noise. Therefore extracting the ENF from the higher frequencies will come in handy since the fluctuations also occur at the harmonics. Clicking on "Extract Audio ENF" on the GUI will execute the program starting first by a downsampling algorithm, utilizing the MATLAB function "resample", that is implemented to reduce the sampling rate from its original value to 140Hz (a bit higher than twice the 60Hz component, to respect the Nyquist sampling criteria).

Furthermore, a narrowband zoom or downsampling to 0.2Hz (0.2 Hz in the case of extracting the ENF from the 60Hz peak and not from its harmonics) around the frequency of interest is performed, to result in a Fourier Transform, using Equation 4.2, applied on the downsampled signal with a Hanning window of 40 samples (200 seconds), yielding a resolution of 5mHz on a 0.2Hz range starting from 59.9Hz to 60.1Hz, at a sampling rate of 0.2Hz. Hanning windows are employed for the FFT analysis of each 200 seconds segment because the "Hanning window is the most commonly used window function for

random signals because it provides good frequency resolution and leakage protection with fair amplitude accuracy" [37].

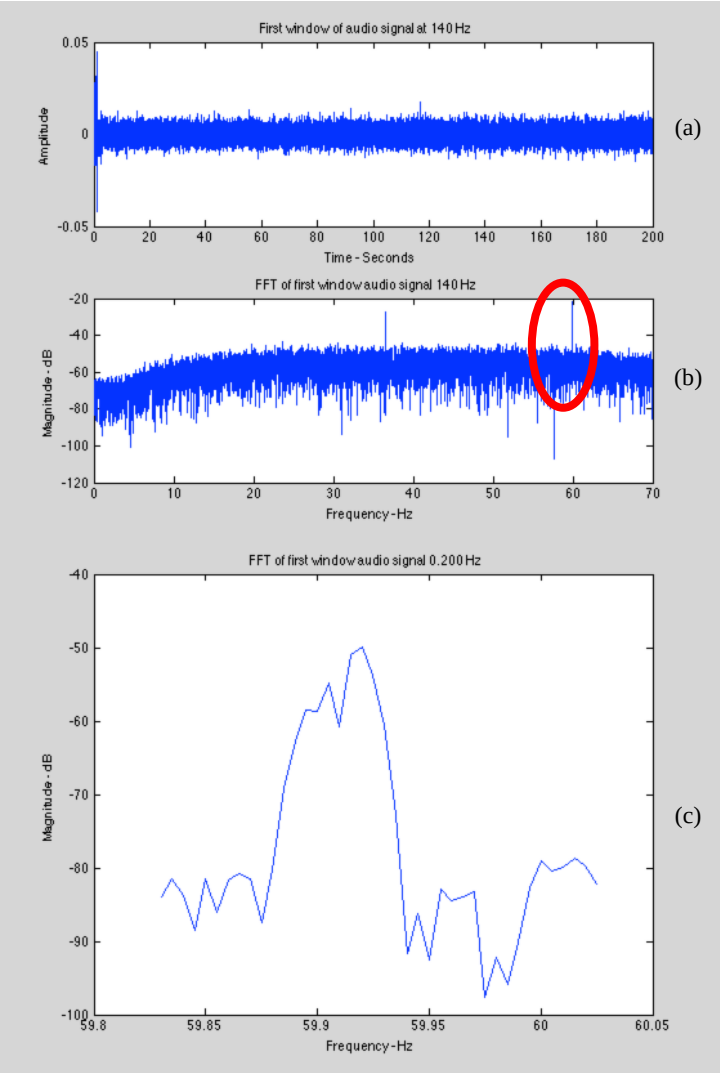

Figure 4.3 – First 200 seconds window of audio a) Time waveform at 140Hz sampling, b) FFT magnitude at 140Hz sampling, c) FFT magnitude of signal around 59.9Hz, signal sampled at 0.2Hz

Figure 4.3 shows the first 200 seconds of audio with its associated FFT magnitude at 140Hz sampling and zoomed around 59.9Hz. In Figure 4.3, the 60Hz peak corresponds to the ENF signal captured with the battery powered Olympus WS210S device. The downsampling process has been implemented on the original signal mainly due to the fact that it would be quite costly (in terms of required FFT size and thus computational complexity and required memory) to obtain an

$$STFT\{x[n]\} = X[\omega, k] = \sum_{n=1}^{L} x[n]w[k-n]e^{-j\omega n} \qquad 4.2$$

accurate value for the exact ENF component located at 60Hz if the original sampling was used, see Figure 4.4b. In order to have a good precision about the exact frequency at

which the 60Hz peak is located and at a reasonable cost, the signal's frequency is therefore shifted by approximately 60Hz (depending on the location of the peak in the FFT magnitude at 140Hz) prior to

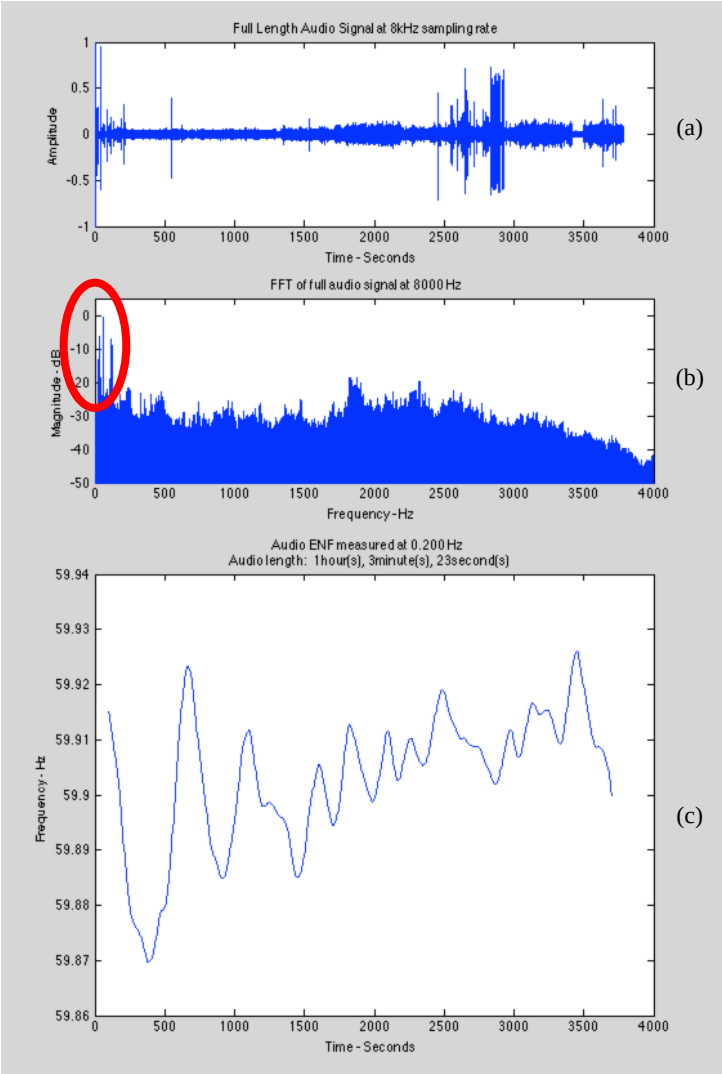

Figure 4.4 – Full audio signal, approximately 1 hour long a) Time waveform at 8kHz sampling, b) FFT magnitude at 8kHz sampling, c) extracted ENF signal using STFT/FFT method

downsampling the frequency rate to 0.2Hz, for a bandwidth of 0.2Hz around 60Hz (i.e. from 59.9Hz to 60.1Hz since the fluctuations do not exceed the previous values). Frequency shifting is performed using Equation 4.3 where " ω_{60Hz_peak} " is the frequency at which the 60Hz peak is located on the 140Hz signal FFT magnitude (see Figure 4.3b).

By analyzing the information presented in Figure 4.3c, a weighted average of the 40 points

$$Signal_{shifted} = Signal_{140Hz} \cdot e^{-i \cdot \omega_{60Hz\_peak} \cdot n} \qquad 4.3$$

spread across the 0.2Hz span can be performed using Equation 4.4, where "$\overline{f_{range}}$" is a 40 points vector ranging from -0.1Hz to 0.1Hz around the peak frequency "ω_{60Hz_peak}" computed previously and "$\overline{|FFT\{signal_{0.2Hz}\}|}$" is the magnitude of the signal's FFT that we divide by its sum to obtain a weighted average result. The corresponding frequency is stored to produce a single ENF sample for the first analyzed window.

$$f_{weighted\_average} = \sum \frac{\overline{|FFT\{signal_{0.2Hz}\}|} \cdot \overline{f_{range}}}{\sum \overline{|FFT\{signal_{0.2Hz}\}|}} \quad 4.4$$

The first window of observed data has a length of 200 seconds. The process described is implemented for each consecutive window subject to a 5 seconds interval shift. This shift value could be increased up to 100 seconds to reduce complexity, but using a smaller shift of 5 seconds allowed a slightly better time resolution and it was convenient because it also corresponds to the window shift used in the AR method, to be described later in Section 4.3, allowing a better comparison and compatibility between extracted ENF signals. For example, this will allow a direct comparison of the reference ENF data extracted from the grid using the STFT/FFT approach or the AR method and ENF data obtained from an audio signal using either the STFT/FFT approach or the AR method. The result of the program execution is depicted in Figure 4.4c where the ENF signal shown corresponds to an audio signal of length 1h:03min:23sec.

Since different devices have slightly different internal clocks, through the sampling process, this leads to slightly different observed ENF frequencies. In other words, recording devices introduce a frequency bias and it can sometimes lead to seemingly falsified outcomes. Table 4.1 presents the frequency bias for the recording devices

utilized for the purpose of this thesis. In the case of the Olympus WS210S device, the bias was centered around 59.89Hz. Searching for ENF component at 60Hz would thus have produced wrong results. For each device, looking for the correct frequency peak maximum around 60Hz was therefore necessary in order to perform the correct frequency shift before applying the downsampling from 140Hz to 0.2Hz, hence leading to the suitable ENF signal.

Recording Device	Frequency Bias
Olympus WS210S	~59.89Hz
Olympus DS150	~59.7Hz
Marantz Recorder	~60Hz

Table 4.1 – Recording devices and their respective frequency bias

The previously described steps are repeated for the amount of windows required to cover the full audio recording. Once the audio ENF signal has been obtained, the user has the option of saving the ENF data as a MATLAB ".mat" file by clicking on "Save audio ENF" on the GUI. The saved information can later be uploaded to the GUI, instead of having to execute the extraction routine all over again, in order to determine an audio file's authenticity.

Utilizing a battery-powered device would require the recorder to capture the 60Hz frequency from radiating electromagnetic waves originating from nearby electric components, refer to section 2.4 for more details. However when a device is plugged straight into the power outlet, the 60Hz peak magnitude greatly increases which definitely yields a better ENF extraction. Table 4.2 compares the relative magnitude of the 60Hz peak of the three recording devices employed, when being battery-powered or mains-powered mode.

Recording Device	Battery-Powered 60Hz Magnitude	Mains-Powered 60Hz Magnitude
Olympus WS210S	-22dB	N/A
Olympus DS150	-25dB	0dB
Marantz Recorder	-58dB	-39dB

Table 4.2 – Magnitude of 60Hz for three recording devices while being battery-powered or mains-powered

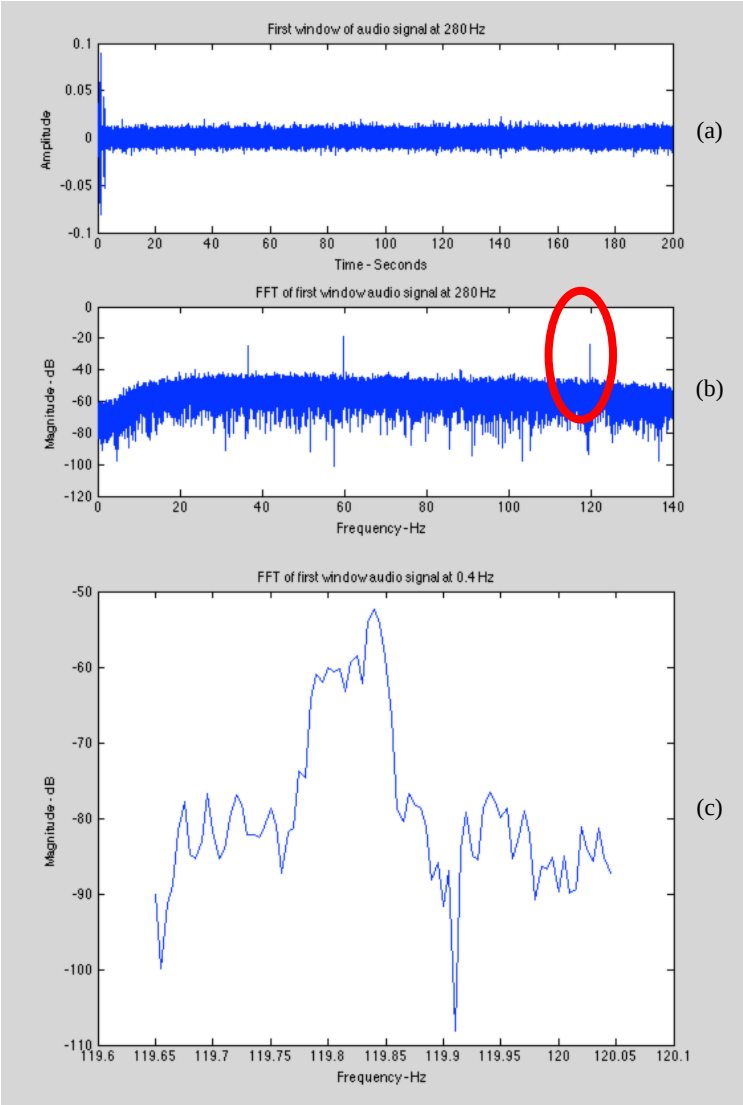

Figure 4.5 – a) First window of audio signal sampled at 280Hz, b) FFT magnitude at 280Hz, c) FFT magnitude around 119.85Hz sampled at 0.4Hz

While these measurements were not formally calibrated (they are relative to the maximum range of the A/D converter), they illustrate the fluctuations between the different models and modes of operation. The values for the Olympus DS150 illustrate the large difference (25dB) between the battery-powered mode and the mains-powered mode for this model. While being mains-powered, the Marantz recorder has a 60Hz peak magnitude that is quite lower (39dB) than the one obtained with the Olympus DS150. The main reason behind such a result is that Marantz

device is well shielded from undesired noise signals, and therefore it does dot capture a significant 60Hz frequency component.

In case where the 60Hz ENF component would be covered with noise, would be attenuated (i.e. audio compression or frequency shielding) or even be erased by a "criminal", the user might alternatively decide to extract the audio file's ENF component from the harmonics situated at 120Hz and 180Hz, if the ENF component at those frequencies are in better condition that the one at 60Hz. By selecting 120Hz or 180Hz from the audio options on the GUI, the algorithm would then first downsample the signal from its original sampling rate (i.e. 44.1kHz or 8kHz) to either 280Hz or 420Hz respectively, before performing a frequency shift and further downsampling around the center frequency of interest.

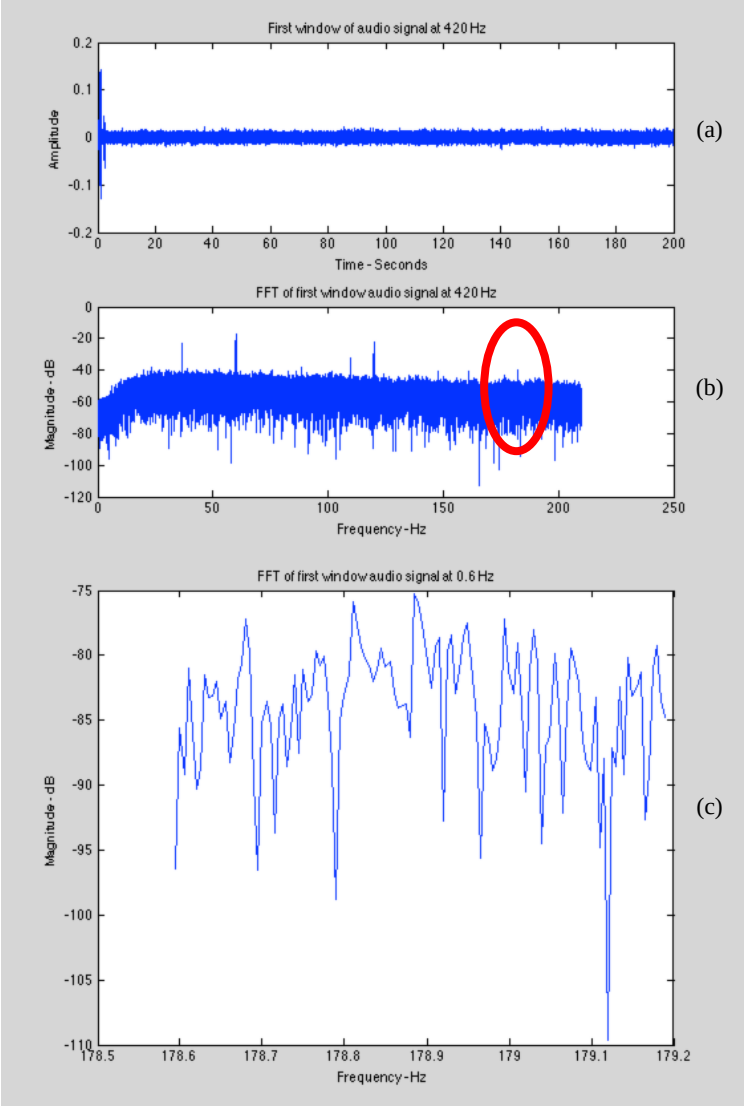

Figure 4.6 – a) First window of audio signal sampled at 420Hz, b) FFT magnitude at 420Hz, c) FFT magnitude around 178.9Hz sampled at 0.6Hz

The peak at 120Hz in Figure 4.5b corresponds to the ENF component captured by the recording device (Olympus WS210S) and the bandwidth of the ENF in Figure 4.5c has doubled to 0.4Hz due to the fact that the search for the ENF component is now at 120Hz instead of 60Hz. The peak observed and the overall shape of the magnitude, are quite comparable to the results found previously at the sampling rate of 0.2Hz. On the other hand, Figure 4.6 depicts the results for the extraction of the 180Hz ENF component. Extracting the ENF fluctuations from the harmonic located at 180Hz means that the bandwidth tripled from 0.2Hz to 0.6Hz. In addition, the peak's shape is very different that what was found earlier for the case of 60Hz and 120Hz, because the audio signal used is mainly a speech recording with stronger speech components mixed with the ENF signal at 180Hz.

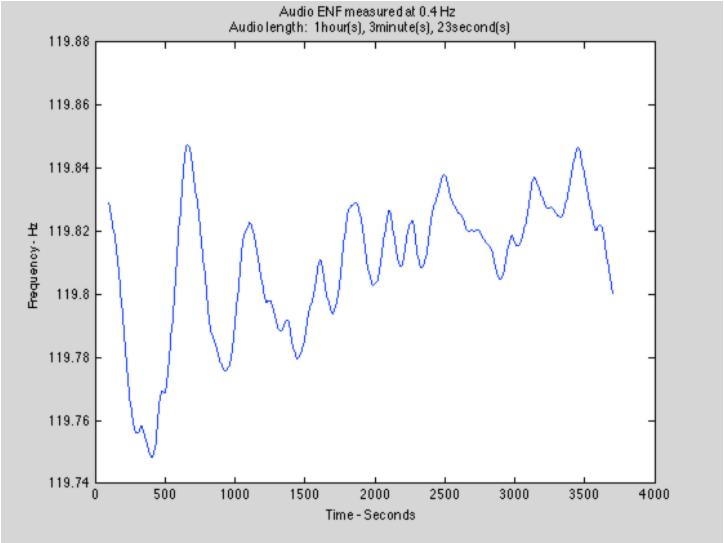

Figure 4.7 – Extracted ENF signal with STFT/FFT method, using 120Hz peak information

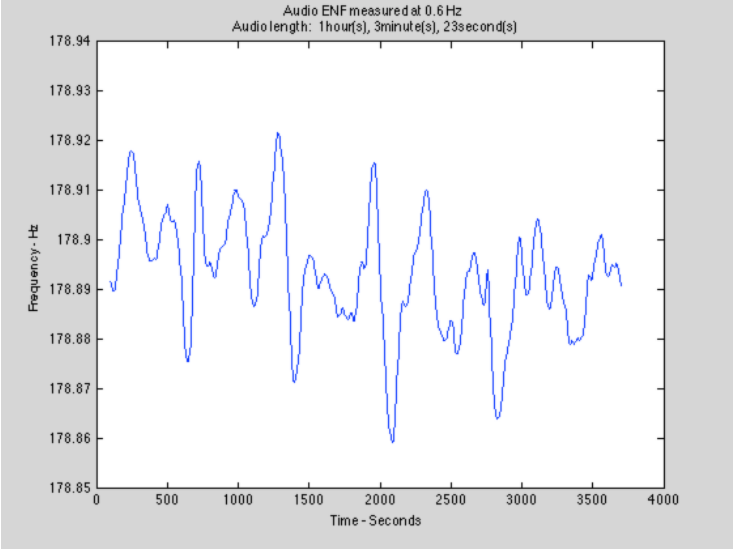

Figure 4.8 – Extracted ENF signal with STFT/FFT method, using 180Hz peak information

Thus based on the ENF to “noise” ratio (where the noise includes all the other sources), the correct ENF signal extraction can be challenging in some cases.

The complete ENF signals extracted from the 120Hz peak and the 180Hz peak are shown in Figure 4.7 and Figure 4.8, by following the same steps as the ones described earlier for the 60Hz scenario (i.e. 200 seconds window with a 5 seconds shift for each computation).

By comparing all three ENF signals computed (Figure 4.4c, Figure 4.7 and Figure 4.8), it is clear that there is a very good match between the ENF extracted at 60Hz and 120Hz, but the ENF component at the 180Hz peak is falsified. The speech levels at 180Hz were too strong, causing the disappearance of the valuable ENF information.

4.1.2. Grid file

The electrical grid signal, which is used to create the ENF reference database and serves as the comparison element later on, has to be captured via the ENF probe explained in Section 2.3. The ENF extraction will begin when the user loads the “.raw” grid file onto the GUI using the “Select Grid ‘.raw’ File” button and then clicks on “Extract Grid ENF” while choosing the appropriate method (e.g., STFT or AR). In order to avoid memory overload, due to the length of the grid file (i.e. ~12 hours at 44.1kHz), the process of reading the sampling from the file is performed iteratively using blocks of 30 minutes, followed by a downsampling algorithm implemented in order to reduce the sampling rate from its original value of 44.1kHz to 140Hz (unlike the audio files where the downsampling process is performed with the full-length signal). To avoid inconsistent outcomes such as unwanted distortions or cut-offs, it is necessary to pre-filter separately every 30 minutes block, prior to downsampling, where the initial conditions of every

consecutive filter are the output results of the previous corresponding filter (however the first filter's initial conditions are set to zero). The original filter created by the function "resample" in MATLAB was modified in a way to reduce its original size from 6301 to 631 (i.e. approximately 10 times smaller). Such a decrease in filter size affects enormously the computational time required to process the 12 hours grid file, thus decreasing the solving time from approximately 25 minutes to approximately 8 minutes while keeping the structural integrity of the grid ENF signal. Figure 4.9 shows the frequency response of the anti-aliasing filter for different filter lengths. The reduction of the filter size and the use of the highly non-ideal frequency response of Figure 4.9c was possible for the filtering of the grid signal, because the 60 Hz component is strongly dominant in the grid file and the resulting downsampled signal does not significantly suffer from frequency aliasing

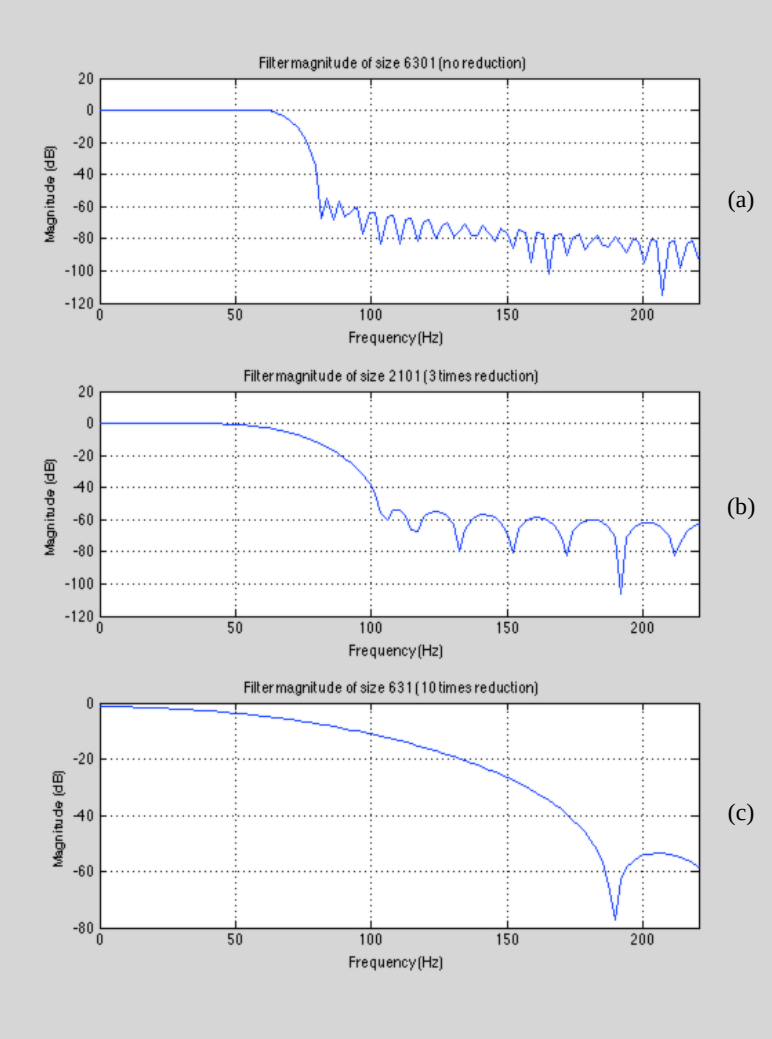

Figure 4.9 – Comparison between the resampling filter subject to different length, a) filter of size 6301 (no reduction applied), b) filter of size 2101 (3 times reduction applied), c) filter of size 631 (10 times reduction applied)

effects. The use of this shorter anti-aliasing filter would not be possible for processing audio files, because the 60 Hz component no longer dominates the rest of the spectrum as in the case of the grid signal.

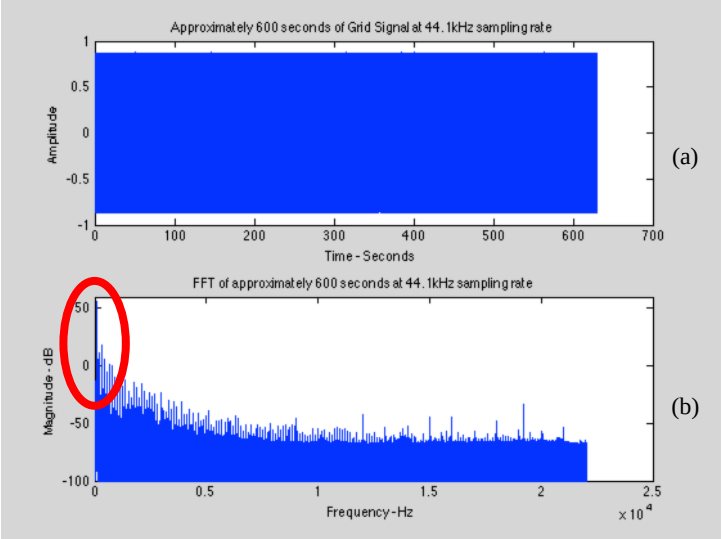

Figure 4.10 – Approximately 600 seconds of grid data, a) time waveform sampled at 44.1kHz, b) FFT magnitude sampled at 44.1kHz

Once again the downsampling process is implemented on the original signal mainly due to the fact that it would be quite costly (in terms of FFT size and thus computational complexity and required memory) to obtain an accurate value for the ENF component located at 60Hz, refer to Figure 4.10b. On the other hand, the 140Hz-resampled signal presents a crisp and clean peak around 60Hz when computing the signal's FFT, as seen in Figure 4.11b. Unlike the audio signal whose signal might most preferably be consisting of speech, hence contaminating the ENF components, the grid signal is a crisp and clean 60Hz signal where the nearby frequencies are significantly

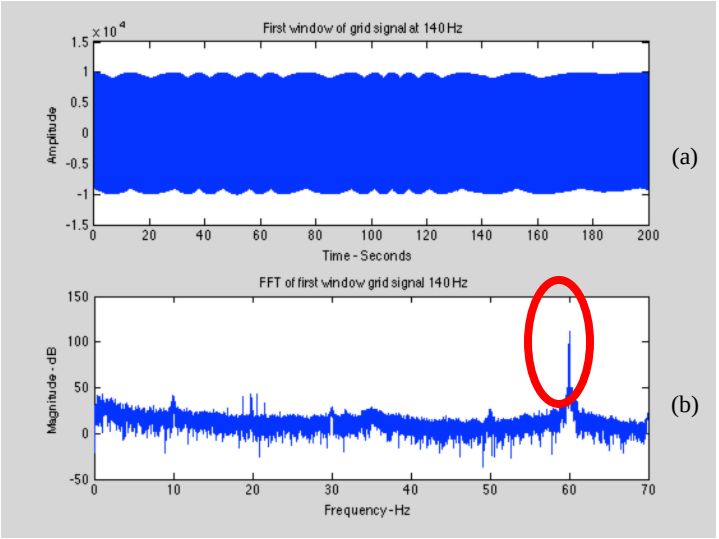

Figure 4.11 – First 200 seconds window of grid data, a) time waveform sampled at 140Hz, b) FFT magnitude of signal at 140Hz

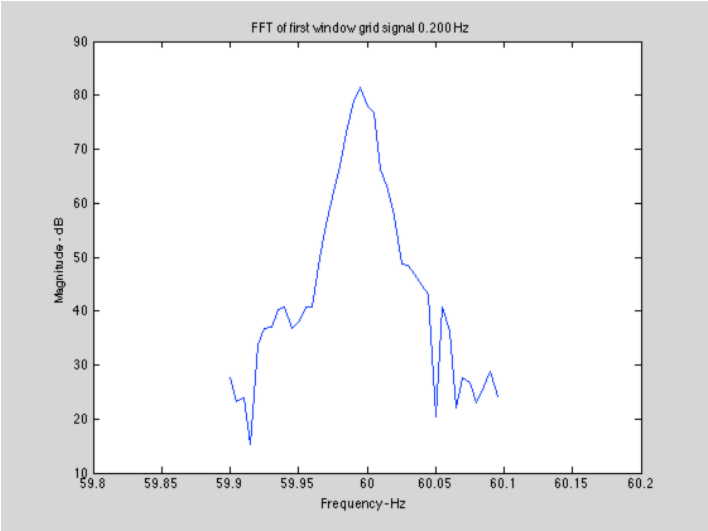

Figure 4.12 – First window FFT magnitude of grid signal around 60Hz sampled at 0.2Hz

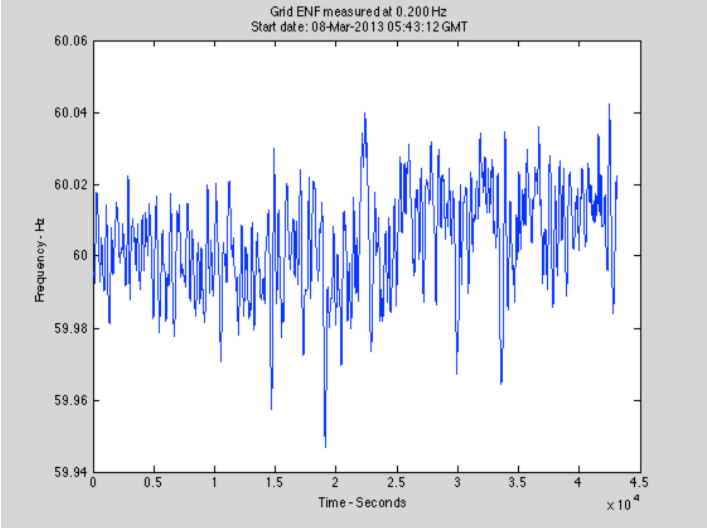

Figure 4.13 – Extracted ENF signal for full 12 hours grid data using STFT/FFT method

low. Therefore the grid ENF will always be extracted from the fundamental peak of 60Hz, as the harmonics in this case are not required.

The complete 140Hz signal is shifted by the frequency at which the maximum of the peak is located followed by an additional downsampling algorithm that reduces the sampling rate to 0.2Hz, for a bandwidth of 0.2Hz around 60Hz. Next a Fourier Transform is applied to a window of 200 seconds (equivalent to 40 samples at 0.2Hz sampling rate) so as to yield a resolution of 5mHz. The next windows are obtained by using a shift of 5 seconds (1 sample at 0.2Hz), as previously described.

The ENF samples corresponding to every window are computed via the weighted average approach defined in Section 4.1.1 and utilizing Equation 4.4 on the 0.2Hz range of the downsampled signal depicted in Figure 4.12. The resulting grid ENF signal, Figure

4.13, is built from consecutive ENF samples. The user is able to save the ENF information as a ".mat" file by clicking on the "Save grid ENF" button on the GUI. Such an option provides the user with the ability to compute the ENF only once for each 12 hours grid file, hence eliminating the need to re-run the algorithm.

At a first glance it is quite difficult to compare the audio's ENF, Figure 4.4c, to the grid's ENF, Figure 4.13, simply by looking at the ENF waveforms. Visually searching for a 1 hour long signal through a 12 hours reference can be quite challenging and exhausting. For this reason, Section 6.1 tackles the subject of comparing the audio's ENF with the grid's ENF by utilizing a sliding window correlation method. The result contains enough information that allows the user to determine the authenticity of the recording, as well as the best possible match throughout the grid ENF reference.

4.2. Zero-Crossing method

The second method to extract the ENF signal was proposed by Catalin Grigoras in [1, 17, 19-21]. The zero-crossing approach is based on determining the ENF fluctuations from the zero crosses of the original grid reference or audio recording. Unlike the STFT method, the zero-crossing approach does not require the conversion from time domain to frequency domain. However working in time domain can certainly have its own limitations and consequences. First the algorithm has to operate at high sampling frequency (i.e. preferably the original 44.1kHz for better precision but at least 8kHz). This is costly and it is the reason why it is mainly utilized to extract ENF for shorter signals. It does however function properly for longer signals but will require much more memory usage and much higher computational time. By comparison, the STFT method functions at a downsampled sampling rate of 0.2Hz, which greatly reduces the number of

samples analyzed. Second, it was reported in [19] that the results of the zero-crossing method are sensitive and can vary significantly depending on the analysis window size, window overlap, etc. (due to the fact that the zero-crossing method is highly nonlinear). For these reasons, for aspects related to the analysis of shorter audio signals in this thesis we are proposing a new method that could work at the reduced sampling rate of 0.2Hz, see Section 4.3, but the zero-crossing approach was implemented nevertheless for testing purposes.

With the zero-crossing method, the need of pre-filtering the signal around the reference frequency mark of 60Hz is mandatory in order to isolate the ENF component and eliminate nearby information. Since we needed a very selective band-pass filter with a very long impulse response, we found that several design methods were not suitable in our practical implementation for such long filters, because they suffered from numerical

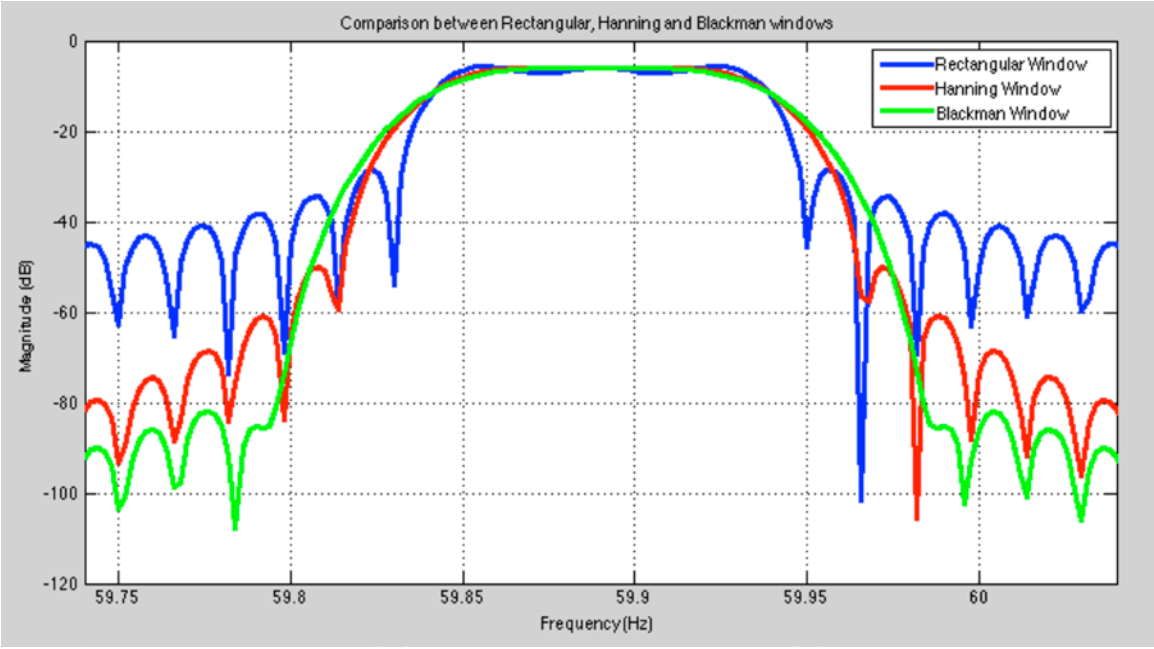

Figure 4.14 – Comparison of side lobes vs. main lobe between different windows functions (i.e. Rectangular, Hanning and Blackman windows)

issues (IIR design methods, FIR design methods based on Parks-McClellan equiripple design, etc.). We thus decided to use a simple analytical window-based FIR design method, which does not suffer from numerical issues. The filter can be constructed via multiple windows (i.e. Rectangular, Hanning, Hamming, etc.) but the Blackman window was found to be suitable in our design. Figure 4.14 depicts the effect of attenuation in the side lobes and the change in the main lobe's width with different windows functions. Clearly the Rectangular window yields to the smallest main lobe, however the Blackman window gives a strong side lobes attenuation, which in fact is the main objective of filtering the signal from 59.9Hz to 60.1Hz if the exact reference mark is 60Hz. The band-pass filter can be constructed directly in MATLAB by simply specifying the lower and higher cut-off frequencies or it can be constructed by modulation (i.e.

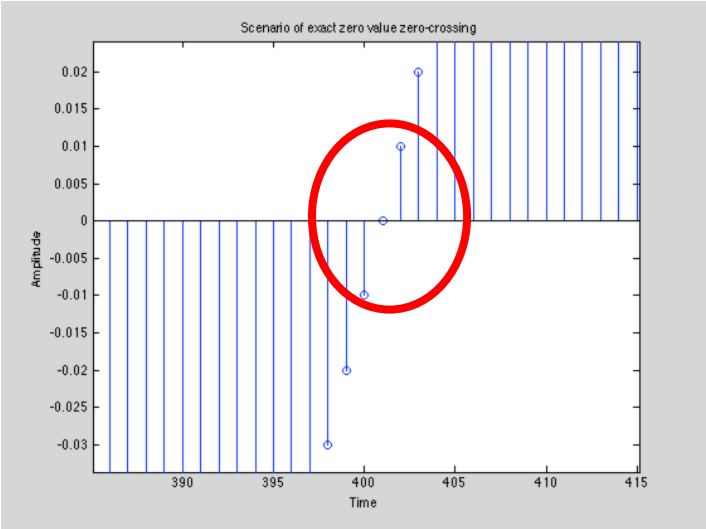

Figure 4.15 – Exact zero value of sample corresponds to zero-crossing

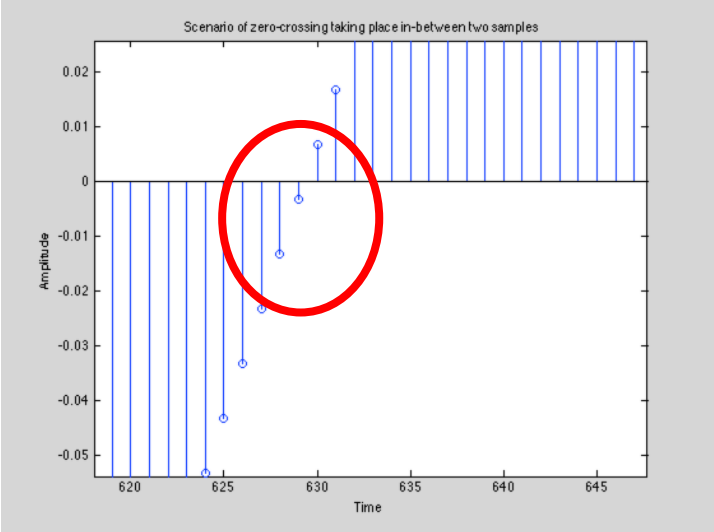

Figure 4.16 – Zero-crossing takes place in-between two consecutive samples

modulation using the cosine function around the frequency of interest) of a low-pass prototype filter.

The advantage of working with a Finite Impulse Response (FIR) filter is that the group delay of the filter is constant, meaning that the filter is considered to be linear phase. The filtered signal will consequently be delayed by half of the filter's size and will affect all frequencies equally. The correct output of the filter can therefore be reconstructed by removing the group delay. From here, the list of zero crosses can be computed in two separate steps. First, the search for the exact zero values, for example in Figure 4.15, must be applied and all matches must return the index at which the match is present. The index is later used to identify the exact time, in the signal, at which the crossing occurred. Second, the algorithm must search for the zero crosses that happened in-between two consecutive samples, as seen in Figure 4.16. If the scalar multiplication of two consecutive samples results in a negative value, this means that a zero-crossing occurred between the two samples and the index of the negative calculation is saved for future reference.

The exact time of the crossing between two samples is determined via a linear interpolation technique, defined in Equation 4.5 where "b" is the value of the first of the two samples analyzed, "x_0" is time at which the first sample is located and $\frac{(y_2 - y_1)}{(x_2 - x_1)}$ is the slope between both samples and "x" is the time of the interpolated zero-crossing.

$$x = x_0 + (-b)\frac{(x_2 - x_1)}{(y_2 - y_1)} \qquad 4.5$$

The difference of time between two consecutive zero-crossings is equivalent to half of the signal's oscillations, therefore the ENF component for every zero-crossing is computed by taking half of the inverse of the time difference between two consecutive zero-crossing, see Equation 4.6.

$$ENF = \frac{1}{2(\tau_2 - \tau_1)} \qquad 4.6$$

4.2.1. Digital audio recording file

Intended for shorter digital audio recordings, in the zero-crossing method implementation we bring the sampling rate of the audio signal to 8kHz (if not originally

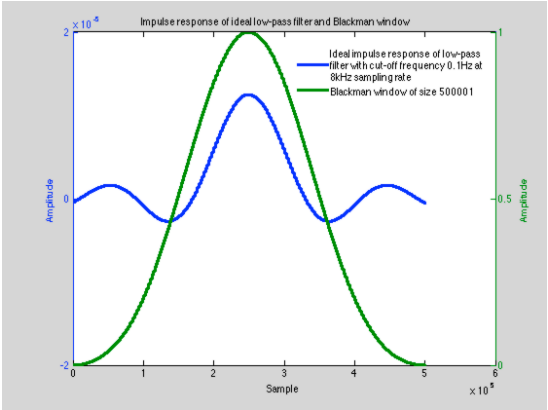

Figure 4.17 – Impulse response of ideal low-pass filter with cut-off frequency 0.1Hz at 8kHz sampling rate (Blue) and Blackman window of size 500001 (Green)

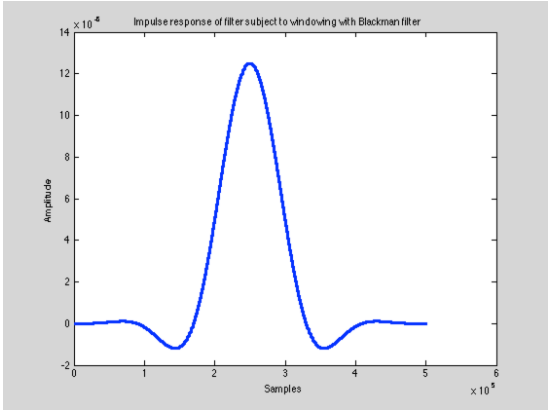

Figure 4.18 – Low-pass filter with cut-off frequency 0.1Hz at 8kHz obtained by multiplying the ideal low-pass filter with the Blackman window of size 500000

sampled at 8kHz) for further analysis. For frequency-biased devices, such as the Olympus DS150 and Olympus WS210S, the Blackman window based filter of size 500000 is constructed having a pass-band range of ±0.1Hz (i.e. 0.2Hz bandwidth) around the bias. The coefficients of the FIR filter impulse response are equivalent to the coefficients obtained from the multiplication of the ideal filter with the Blackman window of size 500000.

A window of that size is necessary in order to obtain a very steep transition band thus preserving the signal in the pass-band and removing more unwanted signal in the stop-band. This was found to be critical for the good performance of the zero-crossing method.

$$w(n) = 0.42 - 0.5\cos(\frac{2\pi n}{M-1}) + 0.08\cos(\frac{4\pi n}{M-1}) \qquad 4.7$$

$$h_d(n) = \frac{\sin(\omega_c(n - \frac{M-1}{2}))}{\pi(n - \frac{M-1}{2})} \qquad 4.8$$

Equation 4.7 is the definition of the Blackman window of size $M-1$ and the ideal low-pass filter impulse response with cut-off frequency ω_c is defined in Equation 4.8. Generally $h_d(n)$ is of infinite length but has to be truncated as to give a FIR filter of length *M*, see Figure 4.17 for visual explanation. Combining the previous two formulas in Equation 4.9 will result in a low-pass filter with Blackman window specifications.

$$h(n) = h_d(n) \cdot w(n) \qquad 4.9$$

The resulting filter's impulse response $h(n)$ of size 500000 is depicted in Figure 4.18 and is modulated around the frequency of interest (i.e. 60Hz or frequency bias), see Equation 4.10, in order to produce a band-pass filter previously depicted in Figure 4.14 for the Olympus WS210S device whose frequency bias is located at ~59.89Hz.

$$h_{\text{mod}}(n) = h(n) \cdot \cos(\frac{f_{bias}}{f_s} 2\pi(n - \frac{M}{2})) \qquad 0 \le n \le M \qquad 4.10$$

The digital audio recording must now pass through the $h_{\text{mod}}(n)$ filter as to produce a clean output with a frequency around 60Hz.

The filtered output is depicted in the time domain in Figure 4.19 and a "zoom" on a range of 200 milliseconds was performed in order to better visualize the filtering effect, see Figure 4.20. On the other hand, the filtering effect on the audio signal can also be seen in the frequency domain in Figure 4.21 where the frequencies around the 60Hz have been lowered from around -30dB to roughly -150dB. Working in time domain is crucial for extracting the zero-crossings and applying the steps described previously in Section 4.2. The ENF component for the analyzed 15 minutes digital audio recording is shown in Figure 4.22.

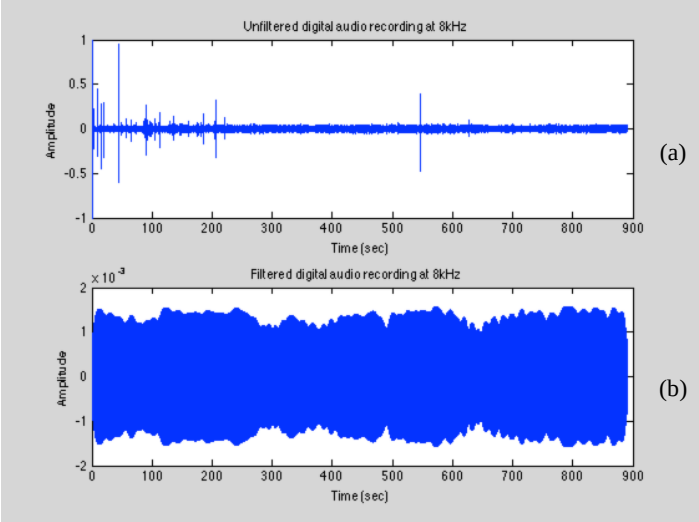

Figure 4.19 – a) Unfiltered audio recording at 8kHz sampling rate, b) filtered audio recording at 8kHz sampling rate

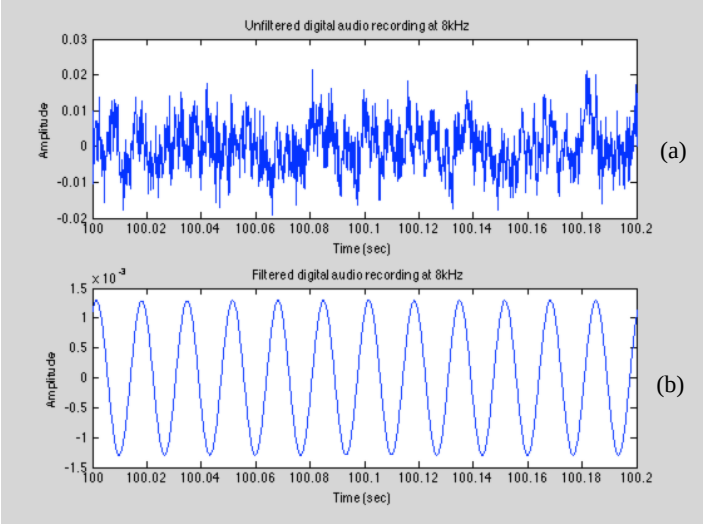

Figure 4.20 – 200 milliseconds of recording from 100 seconds to 100.2 seconds, a) Unfiltered audio at 8kHz sampling rate, b) filtered recording at 8kHz

Like the STFT approach, if the 60Hz component cannot be employed for the ENF extraction, then the possibility of deciphering the harmonics is always available. Extending the search to 120Hz means that the ENF's bandwidth has also doubled and

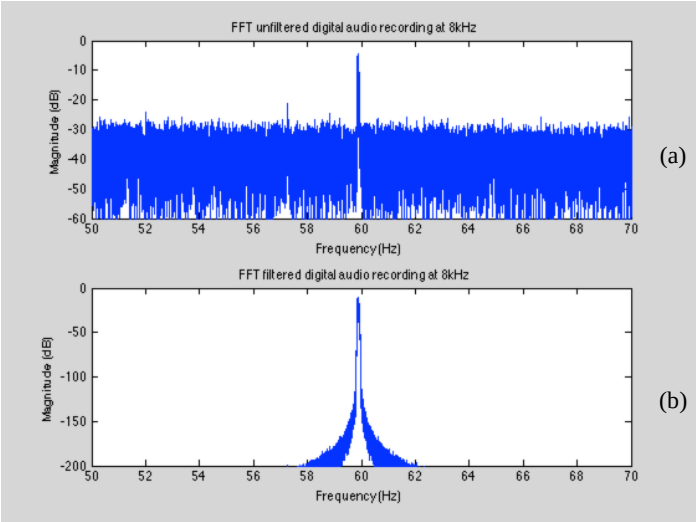

Figure 4.21 – a) FFT of unfiltered recording at 8kHz sampling rate, b) FFT of filtered recording at 8kHz sampling rate around 60Hz peak

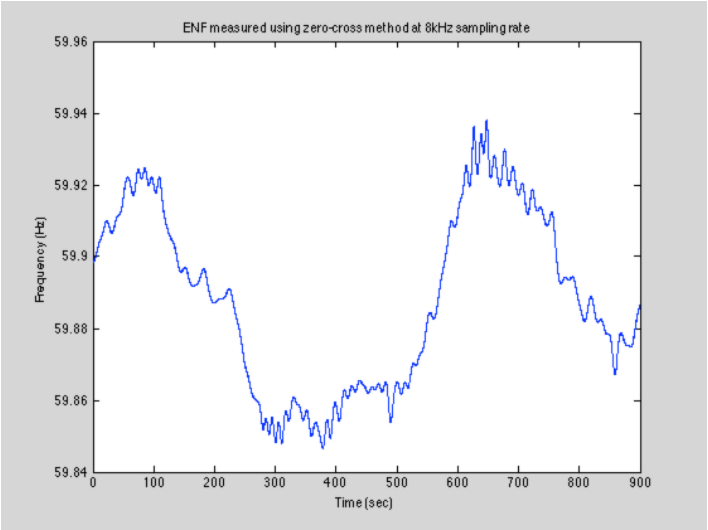

Figure 4.22 – ENF measured with the zero-crossing method for a 15 minutes digital audio file recorded on the Olympus WS210S machine at 8kHz sampling rate extracted from 60Hz peak

now the filtering must be made ±0.2Hz around the 120Hz. Having a larger main lobe means that more unwanted information is included in the filtered signal. In addition, the ENF computation formula utilizing the time of the zero-crossings is modified to Equation 4.11, due to the fact that in one cycle of 120H every zero-crossing occurs at a rate of 240Hz (i.e. twice faster than what is observed when extracting the ENF from 60Hz).

Figure 4.23 presents the 120Hz peak after filtering the signal with a filter of the same length 500000 as employed earlier, but with a larger bandwidth, and it can be seen from the ENF signal in Figure 4.24 that the waveform is practically the same as what was

$$ENF = \frac{1}{4(\tau_2 - \tau_1)} \qquad 4.11$$

computed for the 60Hz peak but is now even noisier. It is evident that by increasing the filter's size the main lobe will be sharper hence unnecessary information will be

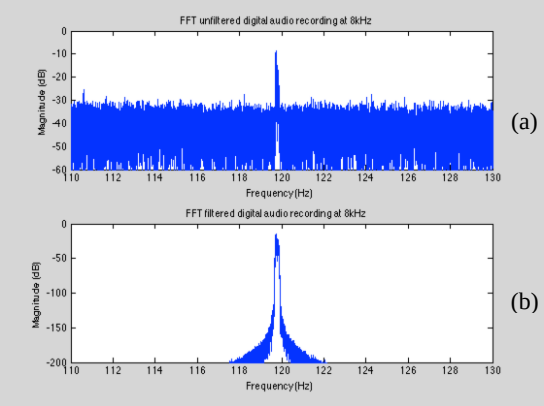

Figure 4.23 – a) FFT of unfiltered recording at 8kHz sampling rate, b) FFT of filtered recording at 8kHz sampling rate around 120Hz peak

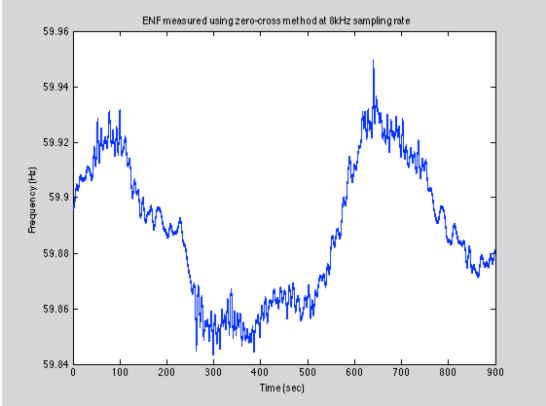

Figure 4.24 - ENF measured around 120Hz with the zero-crossing method for a 15 minutes recording sampled at 8kHz

eliminated but this will yield to more computations requiring more time to extract the ENF signal.

4.2.2. Grid file

Unlike the digital audio files, the reference captured from the probe typically does not suffer from frequency bias (although this depends on the accuracy of the clock and sampling rate of the soundboard) and therefore the ENF components are normally located

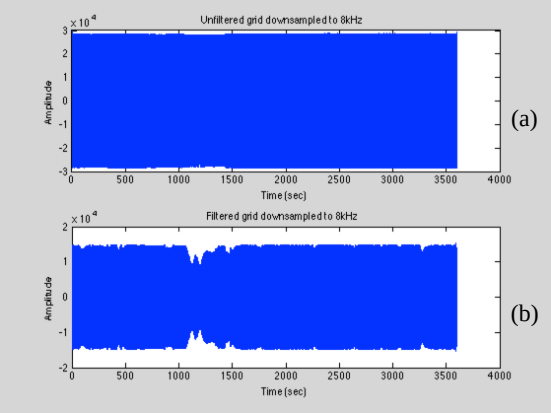

Figure 4.25 – Time waveform of grid signal downsampled from 44.1kHz to 8kHz, a) unfiltered signal, b) filtered signal around 60Hz

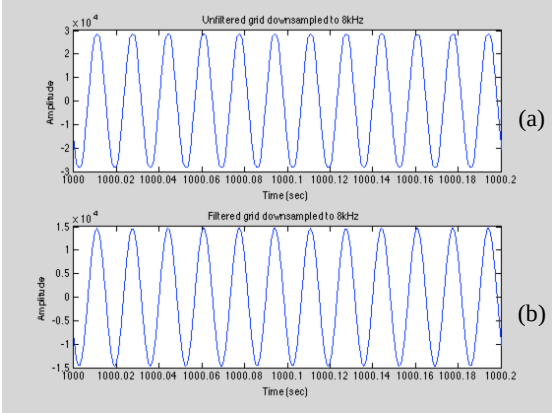

Figure 4.26 – "Zoom-in" on a range of 200 milliseconds, a) unfiltered grid signal time waveform, b) filtered grid signal time waveform

at exactly 60Hz. For simplicity and mathematical reasons (such as ENF comparison between grid and audio), the signal is downsampled from its original rate of 44.1kHz to 8kHz and now matches the sampling rate of the digital audio recording. The band-pass filter having a bandwidth of ±0.1Hz around 60Hz is constructed via modulation of a low-pass filter with Blackman windowing and is depicted in Figure 4.14.

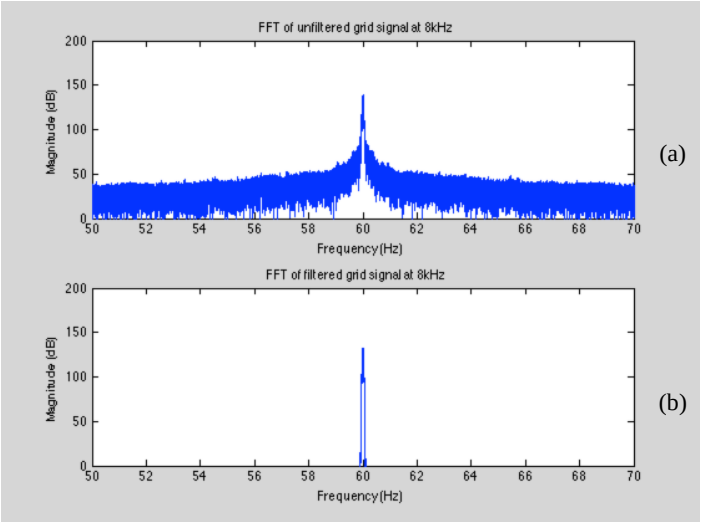

Figure 4.27 – a) FFT of unfiltered grid at 8kHz sampling rate, b) FFT of filtered recording at 8kHz sampling rate around 60Hz peak

It is extremely important to filter the reference data even though the unfiltered signal visually looks like a 60Hz wave in Figure 4.26a. Indeed, this effect can be seen in the frequency domain when plotting the FFT of both the filtered and unfiltered signals. The difference between the 60Hz peak and the noise in the vicinity is in the range of 80dB where as in the audio recording the 60Hz peak was between only 10dB to 30dB above nearby noise. Figure 4.27 shows how the "ENF to noise" ratio can be

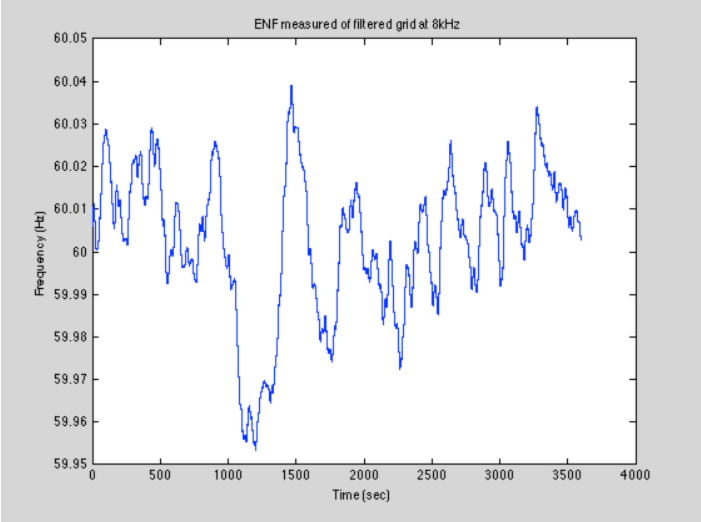

Figure 4.28 –ENF measured around the 60Hz peak of a 1 hour-long filtered grid signal using the zero-crossing method

improved in the vicinity of the 60Hz component in the grid signal, by the application of the bandpass filtering.

Obtaining the zero-crossings and then applying Equation 4.11 will produce the measured ENF for the analyzed file. It is important not to be misguided by the shape of the time waveform when the signal is not filtered (Figure 4.25a), as the algorithm executed on the original signal (i.e. no filtering) yields to a measured ENF that is totally irrelevant and most definitely false, refer to Figure 4.29. A difference of 10^{-5} seconds (cannot be visually noticed) between two consecutive crossings yields to a change of approximately 0.072Hz, which is seen as a major fluctuation in frequency. At the opposite, the extracted ENF from the band-pass filtered signal leads to a correct ENF extracted signal, as shown in Figure 4.28. It is not required to implement the algorithm for extracting the ENF from 120Hz or 180Hz in the grid file since the 60Hz peak will be permanently higher than its harmonics.

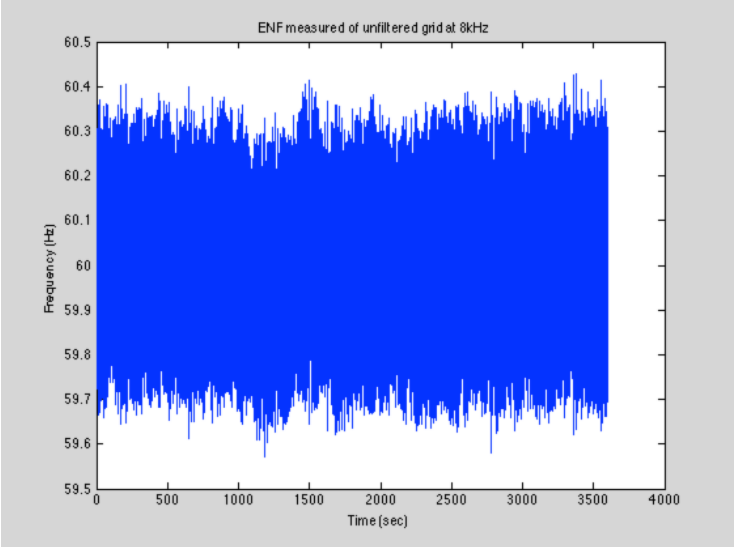

Figure 4.29 – ENF measured around the 60Hz peak of a 1 hour-long unfiltered grid signal using the zero-crossing method

4.3. Autoregressive method

As previously mentioned, this thesis proposes the novel use of a parametric approach using Autoregressive models for ENF extraction. This method is implemented in order to

cover the case of shorter signals (as is the zero-crossing approach) since the STFT approach requires recordings of more than 200 seconds (e.g., obtaining a minimum of 10 ENF points to be used for validation by correlation would require 200 seconds for the first ENF sample and then 5 seconds for each of the other 9 samples, thus a total of 245 seconds). As we will show, the AR method can be used on much shorter signals than the STFT method. Like the STFT method, it is possible to downsample the original signal to 0.2Hz in the case of the AR method, thus leading to low complexity. This aspect is fundamentally different from the zero-crossing method, which is also applicable for short signals but requires an operation at a fairly high sampling rate such as 8 kHz, leading to a high complexity (in addition to being a sensitive, non-linear method).

Figure 4.30 – Basic model for an AR process

A parametric model is defined on the assumption of a particular structure ("model") for the signal being analyzed. This is in opposition to non-parametric methods that do not assume any structure in the signal, e.g. Fourier transforms which are applicable to any signals. Autoregressive parametric models or poles-only models are suitable for signals with peaks or resonance frequencies, thus they are a good fit for the ENF modelling. The parameters or filter coefficients in an AR model can be estimated by finding the coefficients of a filter whose purpose is to predict the current sample of a random process using a finite number of previous samples from the same process. A signal $x[n]$ modeled by an AR process is produced by the output of an all-pole Infinite Impulse Response

(IIR) filter $H(z)$ of order *M*, with a white noise zero mean input $v[n]$, as shown in Figure 4.30, where μ_v is the mean of the input.

The filter response in the z-transform domain is equal to:

$$H(z)=\frac{X(z)}{V(z)}=\frac{1}{1-\sum_{k=1}^{M} w_k^* z^{-k}} \qquad 4.12$$

and equivalently in the time-domain the system is described by the difference equation:

$$x[n]=v[n]+\sum_{k=1}^{M} w_k^* \cdot x[n-k] \qquad 4.13$$

The main objective is to compute the AR model coefficients w_k^* in order to obtain an estimation for the Power Spectrum Density (PSD) of the output $S_{xx}(e^{j\omega})$, which can be directly evaluated using Equation 4.14 knowing the coefficients w_k^* (thus the resulting filter $H(z)$ and its magnitude frequency response $|H(e^{j\omega})|^2$) and determining the variance of the white noise input signal $v[n]$ from Equation 4.15 (thus the spectrum of the signal $S_{vv}(e^{j\omega})$).

$$S_{xx}(e^{j\omega})=S_{vv}(e^{j\omega})\left|H(e^{j\omega})\right|^2 \qquad 4.14$$

$$S_{vv}(e^{j\omega})=r_x(0)-\sum_{k=1}^{M} w_k \cdot r_x(k) \qquad 4.15$$

Then from $S_{xx}(e^{j\omega})$ the ENF frequency can be found by evaluating the peak position in $S_{xx}(e^{j\omega})$. For this application, a model of order $M=1$ was found to be sufficient in order to detect a single peak at the final down-sampled rate. Many methods utilizing a

parametric approach could be implemented (i.e. Yule-Walker, Levinson-Durbin recursion, Welsh-Bartlett, Burg, etc.) but for the purpose of the project only the Yule-Walker approach is developed. This is a classic method first invented in 1907 by George Udny Yule then modified a few decades later by Gilbert Thomas Walker [38].

The algorithm will make use of the observations $x[n]$ to determine the autocorrelation $r_x[n]$ of these observations on a limited sequence of samples (i.e. windows of 4 samples or equivalently 20 seconds) and will be applied to the set of Yule-Walker equations defined in Equations 4.16 and 4.17 [39].

$$\begin{bmatrix} r_x(0) & r_x(1) & \cdots & r_x(M-1) \\ r_x^*(1) & r_x(0) & \cdots & r_x(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_x^*(M-1) & r_x^*(M-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix} = \begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(M) \end{bmatrix} \quad 4.16$$

$$R_x w = r_x \quad 4.17$$

The size of the autocorrelation matrix $\boldsymbol{R_x}$ and the number of AR coefficients represented by the coefficients vector $\boldsymbol{w}$ is dependent on the filter's order M. Therefore, it is important to note that the filter's order M cannot be greater than the number of samples limited by our shifting window. The AR coefficients $w_1, w_2, w_3, ..., w_M$ are obtained by solving Equation 4.18.

$$w = R_x^{-1} r_x \quad 4.18$$

In this specific application the AR model is computed for 20 seconds windows subject to 5 seconds shits. For every analyzed window, the frequency at which the

maximum of the PSD peak occurs is stored and is eventually used to construct the ENF fluctuation signal.

4.3.1. Digital audio recording file

After selecting the appropriate ".wav" audio file as well as the extraction method on the GUI, the algorithm begins by down-sampling the original signal rate to either 140Hz, 280Hz or 420Hz, depending on the peak analyzed for the ENF extraction (i.e. 140Hz sampling rate corresponds to 60Hz peak, 280Hz sampling rate corresponds to 120Hz peak and 420Hz sampling rate corresponds to 180Hz peak). Like the STFT approach, examining the exact maximum frequency peak is important in order to perform the correct frequency shift prior to down-

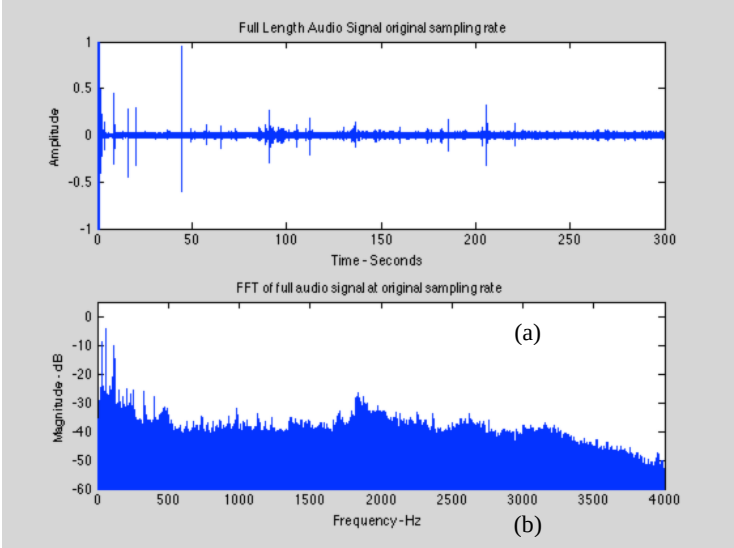

Figure 4.31 - Full 5 minutes audio signal recorded on the Olympus WS210S at 8kHz sampling rate, a) time waveform, b) FFT

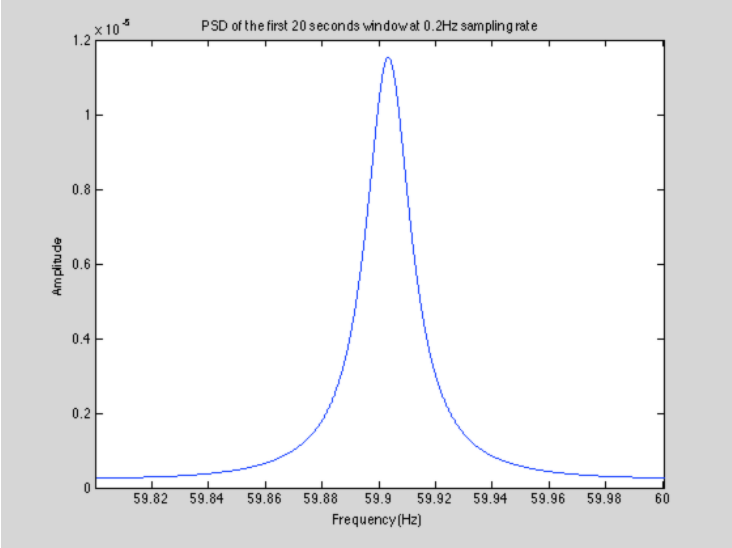

Figure 4.32 – PSD from AR model for first window of 20 seconds of audio signal at 0.2Hz sampling rate

sampling the signal even further to 0.2Hz, 0.4Hz or 0.6Hz, depending on specifications.

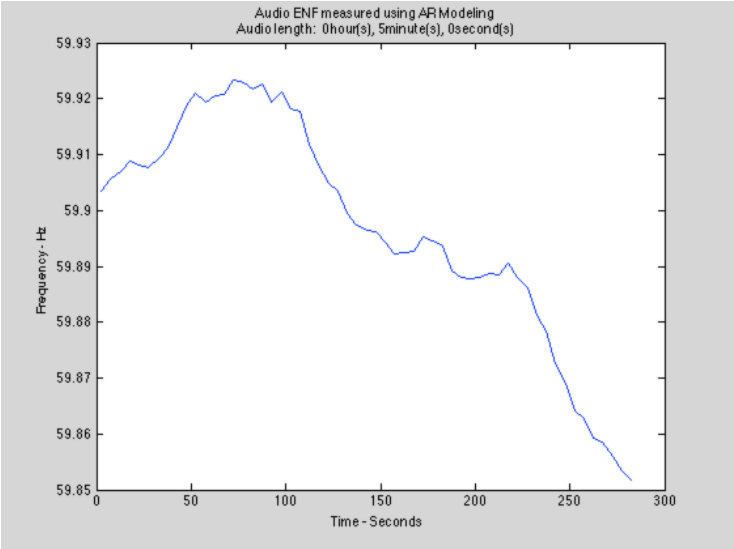

Figure 4.33 – Extracted ENF signal for 5 minutes of audio recording with Olympus WS210S device using the AR approach at the 60Hz peak

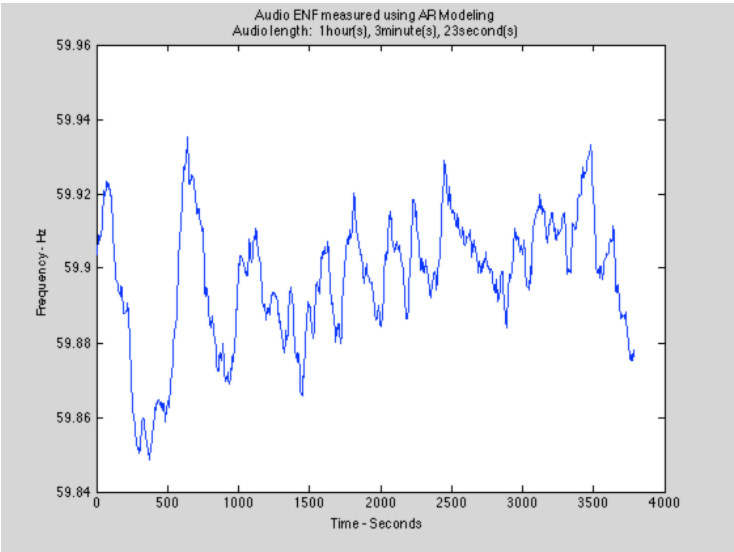

Figure 4.34 – Extracted ENF signal for approximately 1 hour of audio with Olympus WS210S using AR method at the 60Hz peak

Running the algorithm on windows of 20 seconds and applying Equations 4.14 through 4.18 will output the PSD of the model, Figure 4.32, from which the first single ENF frequency sample will be extracted (frequency at which the peak maximum is located). For a 5 minute long signal, the complete ENF fluctuation signal will be comprised of 57 points, as shown in Figure 4.33.

The parametric approach can also be used to extract longer ENF signals (e.g., much longer than 300 seconds), as shown in Figure 4.34, and the user may also decide to extract the ENF information from the peaks located at 120Hz and 180Hz by selecting the appropriate option on the GUI. Although slightly noisier (or, alternatively, showing a better resolution), the ENF signals obtained with the

AR method with the 60 Hz peak in Figure 4.34 and the 120Hz peak in Figure 4.35 clearly have the same overall waveform as the ones obtained previously in Figure 4.4 with the STFT/FFT method.

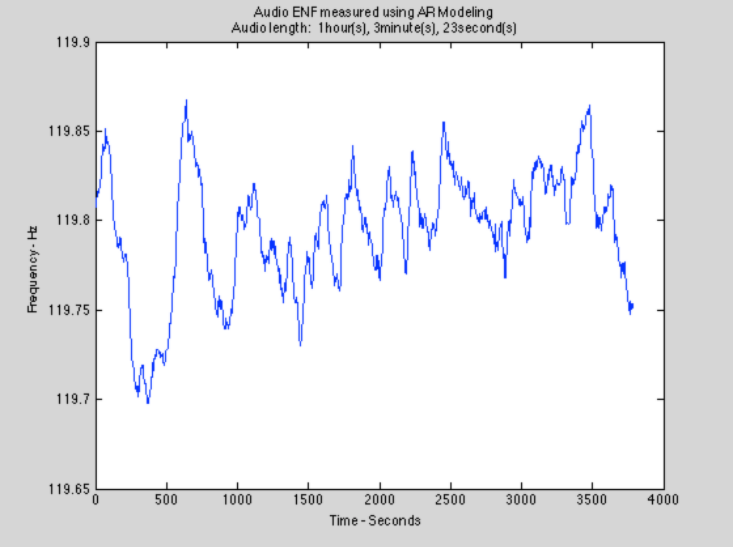

Figure 4.35 - Extracted ENF signal for approximately 1 hour of audio with Olympus WS210S using AR method at the 120Hz peak

4.3.2. Grid file

Unlike the audio recordings, extracting the ENF components from the grid file is only performed on the 60Hz component. In addition, the grid files captured from the probe, see Section 2.3, are saved in blocks of 12 hours and the AR algorithm is capable of handling such size, even though it is originally developed for short recordings. Therefore proceeding with the same steps described for the audio recording file (i.e. down-sampling from original sampling rate to 0.2Hz and then applying the Yule-Walker algorithm), the ENF signal for the reference is calculated and depicted in Figure 4.36.

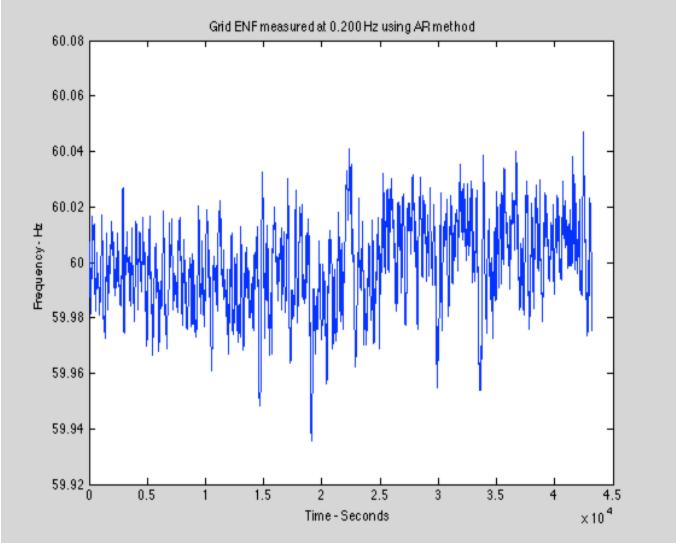

Figure 4.36 – Extracted ENF signal for 12 hours of grid data using the AR approach

The computational time required to run the algorithm is basically the biggest disadvantage of utilizing the AR approach with long files. The extraction time of the ENF depicted in Figure 4.36 was roughly 8-9 minutes. For this reason, it is recommended that the user would save the computed ENF as a ".mat" file, hence not having the need to run the algorithm a second time.

5. Database creation

The creation of the database for the grid ENF signals was implemented using the developed C-program and the probe. The grid files are saved in a ".raw" format along the time-stamps created during the acquisition. The time-stamping is based on the GMT time zone. The probe will continuously capture the electrical signal as long as it is plugged into the power outlet. Over the period of one day the database will record around 7.6GB of information (i.e. each sample has 2 bytes, mono-channel recording, 44.1kHz sampling rate, 3600 seconds/hour), meaning that over one year the user would need a hard drive of 2774GB (or 2.8 TeraBytes), which is available nowadays at a fairly inexpensive price. However the database size can in fact be reduced if the sampling rate is lowered to 8kHz (depending on the soundboard available), as previously mentioned. Table 5.1 illustrates the required storage space for different sampling rates.

Sampling rate	Hard drive size needed for 1 day of data	Hard drive size needed for 1 year of data
44.1kHz	~7.6GB	~2.8TB
32kHz	~5.5GB	~2TB
8kHz	~1.3GB	~0.5TB

Table 5.1 – Hard drive size needed for one year of signal recording, depending on the sampling rate used to capture and save the grid's ENF information

The user can also build his own database of extracted ENF by saving the computed ENF with the GUI. By doing so, he has the option of loading the ".mat" file onto the GUI for testing purposes rather than running the extraction from the large original ".raw" grid file, and the original large ".raw" file may no longer need to be kept on the hard drive, see Figure 5.1 for a comparison of the 3.8GB and 53 KB resulting files.

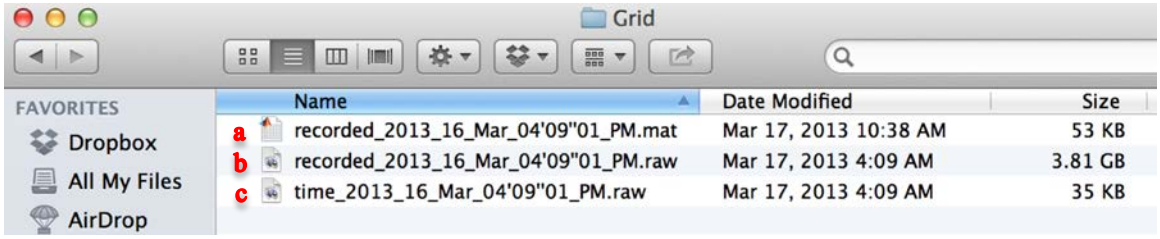

Figure 5.1 – Result of the C-program, a) ".mat" file created after extracting the ENF, b) original ".raw" file containing the ENF data, c) original time-stamp file

Prior to saving the ".mat", the software detects the algorithm used for the ENF extraction then will insert a number in a new cell at the end of the data. For example, a "1" will represent the AR method, a "2" will represent the STFT approach, etc. This step is important for the testing phase, as every method may be different from the other (different ENF resolution due to the extraction algorithm).

For security reasons, it is recommended to capture the grid signal on a computer that is only intended for this specific application. In addition, the computer should not be connected to any Internet connection thus eliminating the risk of obtaining viruses or getting hacked. The computer on every electrical grid must be placed in a secured location where supervision is permanent. Gathering ENF data globally will increase the chance of determining a recording's authenticity, wherever the audio has been made. Grigoras configured the synchronous grid of Continental Europe (i.e. formerly UCTE) [17, 19, 20], while Cooper established the database for the United Kingdom [40-42]. Moreover, the ENF database in the United States is divided into three grids where the Target Forensic Services Lab in Minneapolis studies the Eastern grid and the NCMF gathers the data in the Western grid. However the ENF fluctuations are currently not being gathered yet in the state of Texas.

It is mandatory that all computers be connected to an Uninterruptable Power Supply (UPS) in case power surges occur. In addition, the gathering of ENF data can also be acquired on multiple computers that are run in parallel, hence reducing the risk of missing any samples. As previously mentioned, for the purpose of the thesis the C-program captures data in blocks of 10 seconds, hence giving ample time to write the information from the buffer to the file while the buffer fills up again.

6. Testing

The testing phase consists of determining a recording's authenticity once the ENF signals have been extracted from the grid reference and from the digital audio file. By applying the correlation approach, see Section 6.1, the algorithm can determine where a match occurs between the audio and grid files and whether a discontinuity is present in the audio file, hence labelling the recording as unauthentic. In order for the correlation algorithm to run, both ENF signals must be present. Therefore the user can proceed with the verification steps in two different ways. First, the correlation can be computed by clicking on "Compute Correlation of ENF sequences" on the GUI directly after extracting the ENF information from both files (i.e. it is not required to save the ENFs as ".mat" files in order to compute the correlation). Second, the correlation can be computed by loading the corresponding ".mat" files onto the GUI (i.e. if the user previously extracted the ENF and then saved them as ".mat"), by pressing on "Load Grid ENF Data ".mat"" and "Load Audio ENF Data ".mat"", followed by a button click on "Compute Correlation of ENF Sequences". In fact, the user also has the option of extracting the ENF signal from the audio and then loading the ".mat" file for the grid, or vice-versa, before correlating both signals (i.e. this is very efficient since the ENF extraction from the grid file takes ~8 minutes).

Various tests and results for different scenarios are presented in Section 6.3. Unaltered and altered digital recordings of multiple lengths will be compared with different grid files and the different approaches elaborated in Section 4 will also be compared with each other. In fact, it is important to say that the ENFs extracted using the AR method and the STFT method can easily be correlated together since they both utilize

windows that are subject to 5 seconds shifts (thus leading to a 0.2Hz ENF signal production rate, which is independent on the final sampling rate of the audio file). However, it is preferable and recommended to compare the ENFs of the grid and audio files obtained using the same method of extraction. On the other hand, an audio ENF extracted with the zero crossing method can normally only be compared to a grid ENF extracted using the same approach, due to the difference in the production rate of the ENF signal, i.e., since the ENF samples are produced at a rate corresponding to twice the considered frequency, for example a rate of 120 ENF samples per second if the 60Hz component is considered for extraction.

The Olympus WS210S device is utilized for most of the experiments since it is more modern than the Olympus DS150 and the Marantz recorder. Nonetheless, it will be shown that the two latter devices also embed the ENF component. It is important to note that the capturing of the reference grid was done at the University of Ottawa (electrical outlet from Room 5-130 in the SITE building) and the audio files recorded on the Olympus WS210S device were made in a different location at a distance of about 2km from the former. The results will therefore indicate that indeed the electrical fluctuations occur simultaneously and in concordance throughout the same electric transmission grid, as explained in Section 2.1.

6.1. Cross-correlation approach

A comparison approach based on normalized cross-correlation is applied once the audio recording's ENF and the grid's ENF have been deciphered. Note that we always assume that the grid signal is longer than the audio signal. Correlating both signals using the un-normalized "*xcorr()*" function in MATLAB will determine the audio's ENF

signal's "best match" in the grid's ENF signal. Normalizing the obtained correlation value by the square root of the product of the total energy of both matching segments, see Equation 6.1 and the short MATLAB code below, will produce a normalized correlation value between +1 and -1 [21].

```
[x,lags]=xcorr(GridENF,AudioENF); % x is the correlation here
[max_value,max_position]=max(x);
offset_position=max_position-length(GridENF);
aenergy=sum(AudioENF.*AudioENF); % total energy a
b_segment=GridENF(1+offset_position:length(AudioENF)+offset_position);
benergy=sum(b_segment.*b_segment); % total energy of b
x=x./sqrt(aenergy*benergy);
```

$$\rho_{xy} = \frac{\sum_{n=1}^{N} (audio[n] - \overline{audio}) \cdot (grid[n] - \overline{grid})}{\sigma_{audio}\sigma_{grid}} \qquad 6.1$$

ρ_{xy} defines the normalized cross-correlation coefficient , σ_{audio} is the rms value of the audio ENF signal, σ_{grid} is the rms value of the matching segment analyzed in the grid ENF and N is the length of the audio ENF signal. A strong correlation between the audio ENF and the grid ENF means that the correlation coefficient value

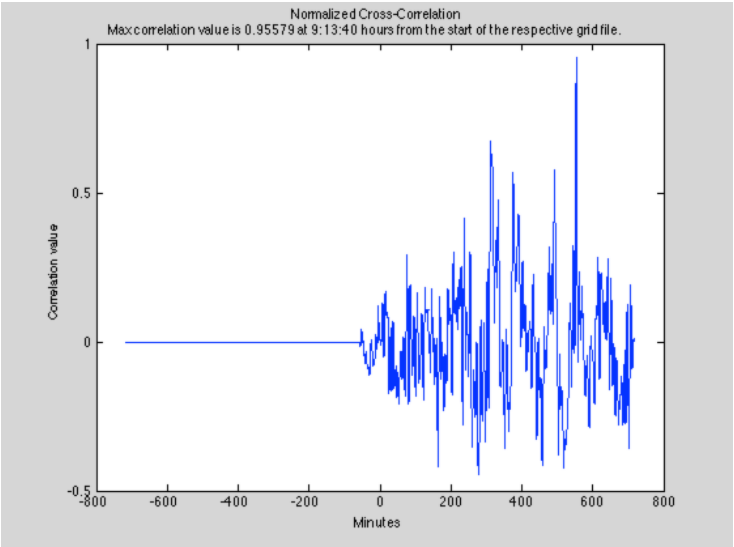

Figure 6.1 – Normalized correlation result with maximum correlation of 0.955 at 9:13:40 hours from the start of the grid file

will be closer to +1, whereas the value of 0 represents no correlation between the signals. In case there is an exact match, the correlation coefficient would be equal to +1. We note that a value of -1 refers to having two identical waveforms where one is horizontally inverted with respect to the other. This being said, for this application of ENF matching the algorithm should look for the maximum positive correlation coefficient value, indicating the location at which the best possible match took place. The result is plotted in a graph where the "y" axis is the correlation coefficient value and the "x" axis represents the time, see Figure 6.1.

The time where the maximum value occurs can be extracted, in hours, minutes and seconds. This information is quite valuable since it can be used to obtain the exact time at which the audio recording has been made. The recording's date and time can be precisely determined by adding the maximum correlation time, found previously, to the time stamp of the grid's ENF signal. If the time stamp file is not found during the execution of the algorithm, then the recording's date of time will be with respect from the start of the grid.

Once the algorithm is done, the result is depicted on the GUI in the white bar beside "Compute Correlation of ENF Sequences", see Figure 6.2.

Figure 6.2 – Demonstration of how the result is presented to the examiner once the correlation is applied

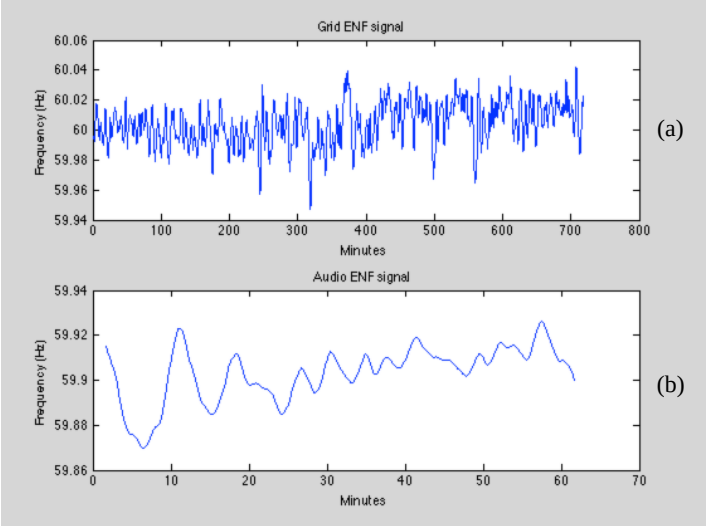

Figure 6.3 – a) Full 12 hours grid ENF signal, b) audio ENF signal

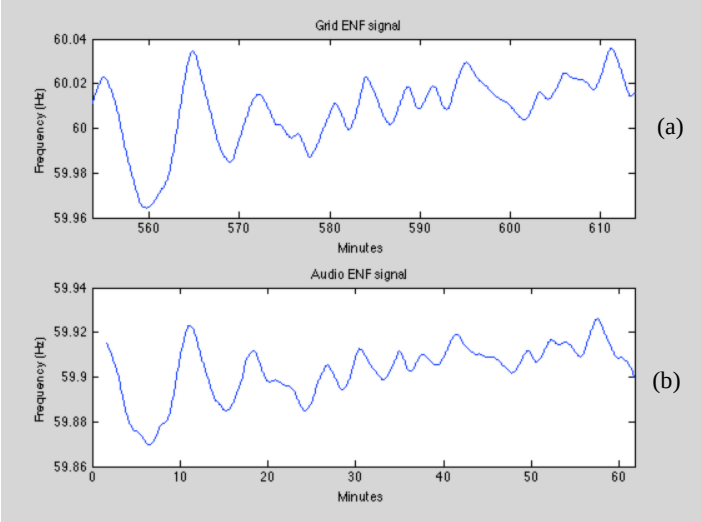

Figure 6.4 – a) Zoom on the grid ENF signal, b) audio ENF signal

Moreover, the grid ENF and the audio ENF are both depicted in the same figure, hence facilitating the visual comparison, see Figure 6.3. The user also has the option of "zooming-in" at the corresponding time by clicking on the "Zoom in" button on the GUI, with the resulting curves shown in Figure 6.4. This option is definitely important because manually searching through a 12 hours reference grid signal can be strenuous.

6.2. ENF discontinuity detection techniques

One of the most common ways of editing digital audio files is by the applying the classic technique of butt-splicing [43, 44], which is based on process of removing a small audio part somewhere within the original digital audio recording and replacing it with another. Cooper states that "when an audio waveform is butt-spliced, the amplitude of the

sample either side of the splice point is unlikely to be in equilibrium producing an abrupt change or 'discontinuity'" [44], see Figure 6.5.

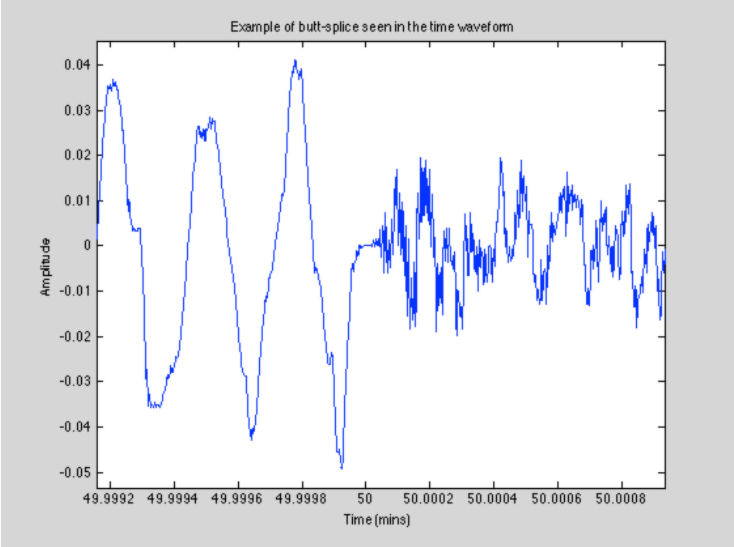

Figure 6.5 – Example of butt-splice where the amplitude of the sample of the original signal is not in concordance with the spliced recording

However, meticulous alterations and manipulations render the signal hard to classify as unauthentic, mostly in the time waveform situation. Even listening to the audio file can sometimes lead to inaudible edits. Automated search algorithm must be developed because visually analyzing unknown recordings is extremely exhausting and practically impossible to detect edits, especially when the sampling rate is high. The user would need to visually verify 158,760,000 samples if the audio file is 1 hour in length and sampled at 44.1kHz. Even for an audio recording that is 5 minutes in length and sampled at 8kHz the user would have to scan 2,400,000 samples.

For the purpose of the thesis, butt-splices will be analyzed on the ENF signal (i.e. grid frequency fluctuation vs. time) instead of the time waveform signal. Correctly butt-splicing a recording in the time domain is feasible, however correctly placing an ENF frequency can be much more laborious and burdensome. Moreover, utilizing different recording devices having different frequency bias will yield to an ENF signal that is "discontinuous" in frequency at the time of the edit.

The first derivative approach is implemented and is intended to detect sudden jumps between two consecutive samples. Once the ENF for the audio signal is computed, the algorithm will apply Equation 6.2 (where $x[n]$ is the ENF extracted from the audio signal, $x'[n-1]$ is the first derivative value (or first difference value) obtained and N is the length of the ENF signal) in order to detect the absolute change in magnitude between two samples.

$$x'[n-1] = |x[n-1] - x[n]| \qquad 1 \le n \le N \qquad 6.2$$

The result of the derivative for the signal in Figure 6.6 is plotted in Figure 6.7, where unusual large peaks represent a big change in ENF value between samples. If the value of the first derivative at any sample exceeds a predefined threshold then the corresponding point in time is labelled as being an edit in the audio recording.

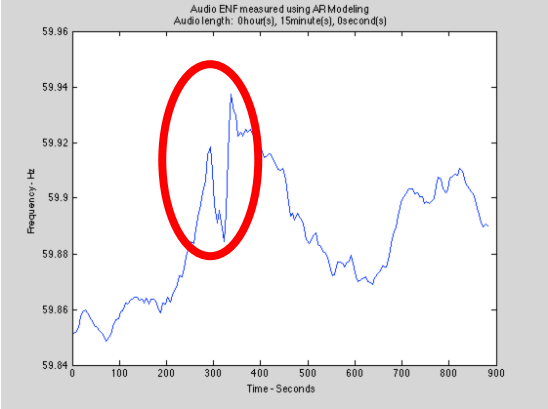

Figure 6.6 – Extracted ENF for a 15 minutes recording that has been tampered with 30 seconds of audio at approximately 300 seconds, both originating from the Olympus WS210S device

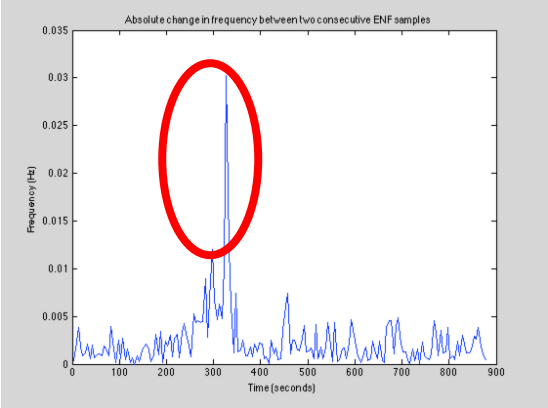

Figure 6.7 – First derivative of the ENF signal from Figure 6.6 where the unusual peak represents a sudden change in frequency at the corresponding time of 300 seconds.

For better visualization and identification purposes, the algorithm will also output the ENF sequence with vertical green lines, located at the points where the threshold is surpassed, and one red line, located at the time where the highest difference between two samples is observed, see Figure 6.8. In addition, the corresponding time for the highest

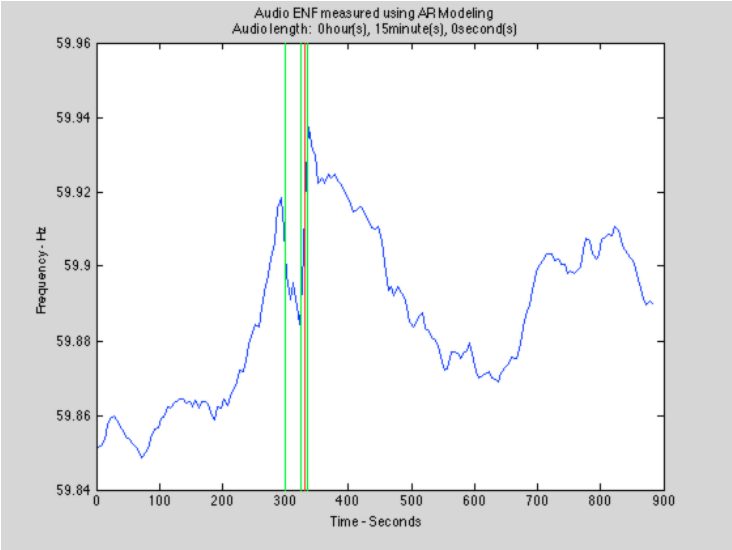

Figure 6.8 – ENF sequence for modified audio with vertical lines indicating probable alterations

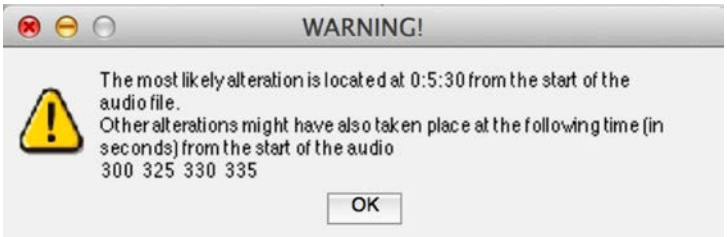

Figure 6.9 – Pop-up message indicating the time of the probable alterations in the extracted ENF signal

frequency change is written in a message bar, see Figure 6.9, whereas the other time indices corresponding to the remaining probable alterations are simply enumerated.

Having different threshold value is required due to multiple extraction methods that can also be performed on different ENF frequencies. Table 6.1 presents the different threshold values that are used in the algorithm. The threshold values have been manually set to a reasonable level after performing multiple butt-splice detection tests for altered and unaltered signals. A satisfying value could ultimately be obtained by having a statistical distribution of data from a large number of tests, see Section 7.2.

Sampling rate	Threshold, STFT method	Threshold, AR method
0.2Hz	0.0021Hz	0.01Hz
0.4Hz	0.0033Hz	0.02Hz
0.6Hz	0.0042Hz	0.025Hz

Table 6.1 – Threshold values employed during the butt-slice detection algorithm

6.3. Various tests and results

6.3.1. Olympus WS210S audio recording device

Unaltered audio recording 60 minutes in length, ENF extraction using STFT method around 60Hz peak

Figure 6.1 shows the correlation results when applying the STFT/FFT ENF extraction algorithm around the 60Hz peak to a 1 hour audio recording. The correlation value of 0.955 is considered to be high, thus indicating that a match of the audio ENF has been found at the time 9:13:40 from the beginning of the grid ENF. In addition, due to the fact that the grid file has been time-stamped (i.e. start date of the grid file is at 05:43:12 GMT on 8-mar-2013) it was possible to determine the exact time and date of the audio recording by adding the maximum correlation time to the time-stamp (i.e. recording started at 14:56:52 GMT on 8-mar-2013). The results as shown on the GUI were presented in Figure 6.2. In case the time-stamp file is not found (it is recommended to keep the time-stamp file of the corresponding recorded grid file in the same folder) the correlation time is presented with respect to the start of the grid, see Figure 6.10.

Since the reference grid is 12 hours in length, the user can have quite some difficulty locating for the match throughout 12 hours of data, as shown in Figure 6.3. The zoom-in option will therefore locate and then plot the grid file at the location of the match for the same duration of the audio file, depicted in Figure 6.4. In this scenario it can be said that the audio recording in question is in fact authentic since the correlation value computed is high and since the audio ENF extracted is nearly identical to the reference grid ENF, which has also been extracted using the STFT approach.

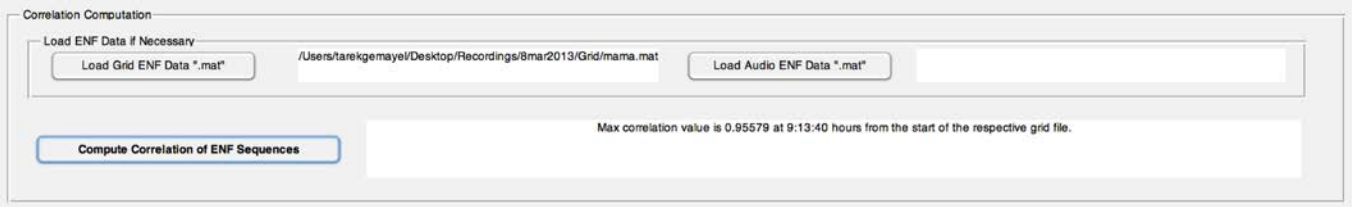

Figure 6.10 – Result output when time-stamp file is not found

Case of an audio signal that was not recorded during the interval of a reference grid file (hence no match is found)

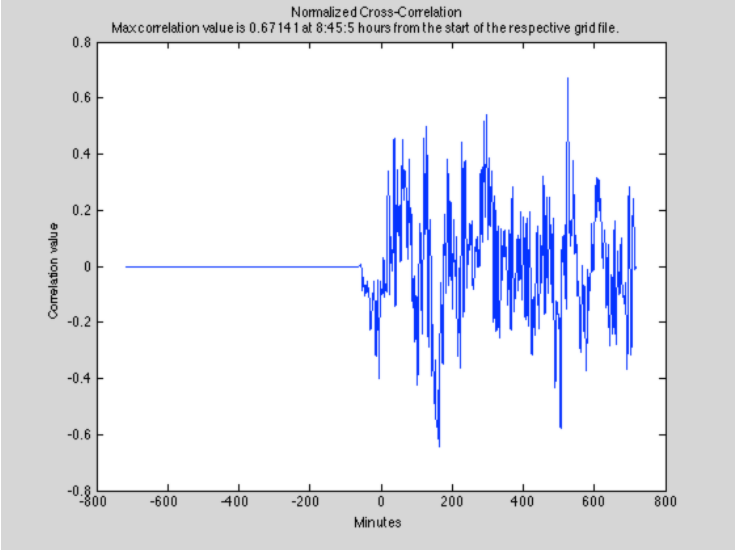

Figure 6.11 – Normalized correlation result between audio file recorded on March 8th and grid file recorded on March 16th

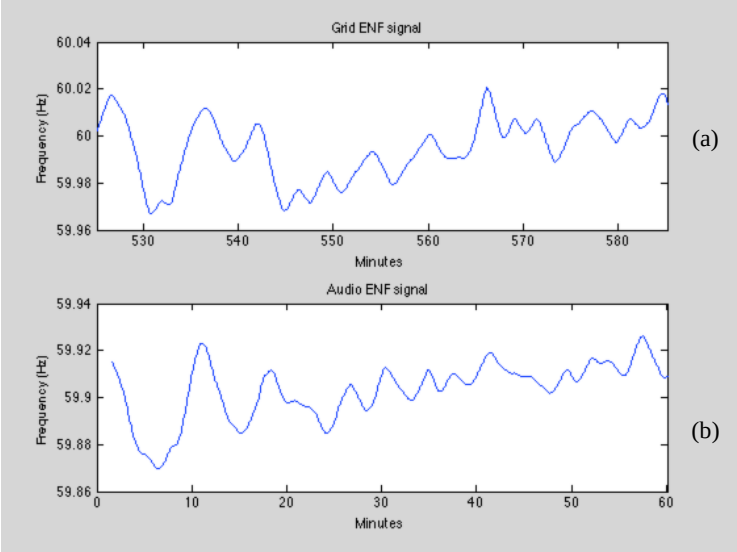

Figure 6.12 – a) Zoom-in at best match found between grid ENF at maximum correlation time and a) audio ENF

If the audio recording's ENF is being compared to a reference grid whose ENF does not contain a match with the recording then the correlation result obtained will be low, as shown in Figure 6.11.

In this scenario, the audio file recorded on the 8th of March 2013 will be correlated with a reference created on the 16th of March 2013. The normalized correlation value is 0.671. The user still has the option of zooming in at the

correlation time (i.e. 8:45:50 hours from the start of the grid), which will help to visually confirm that the audio's ENF is not present in the grid's ENF, as shown in Figure 6.12. Although there are common trend in the two waveforms, a detailed analysis shows significant differences

Audio recording 60 minutes in length with 10 minutes of content modified, ENF extracted using STFT method around 60Hz peak

For a case where 10 minutes of a 60 minutes audio recording has been altered/modified, then computing the recording's ENF should result in waveform having discrepancies at the time of modification. The extracted ENF for such a case is shown in Figure 6.13. In addition, a detection measure was implemented in order to detect sudden changes in ENF signal and will warn the user with a popping window containing the location of most probable alteration as well as the location of the other alterations that may have taken place (see Section 6.2 for more details), as shown in Figure 6.14.

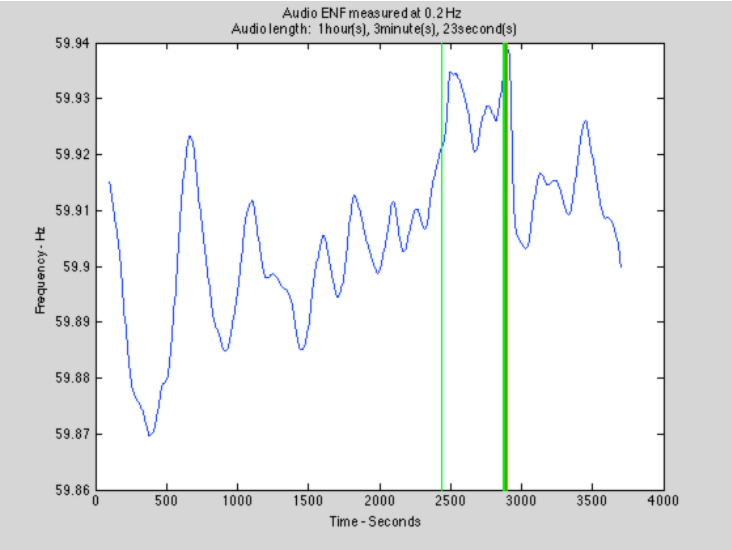

Figure 6.13 – Modified audio ENF at 40 minutes from the beginning of the recording

(a)

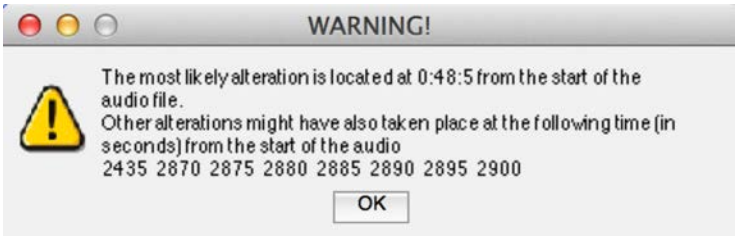

Figure 6.14 – Warning sign indicating the time at which the alterations might have taken place

(b)

Clearly the audio ENF shows that some alteration has taken place, but the warning sign indicates that the most probable alteration has been around 48 minutes from the start of the audio, which is also represented by the red vertical line in Figure 6.13. The correlation of the altered recording with the grid ENF is presented in Figure 6.15 and the comparison of both ENF signals is shown in Figure 6.16. The correlation value obtained is 0.825, which is not as high as when the genuine recording was used, however the maximum correlation time obtained is still the same as before (i.e. 9:13:40 hours from start of grid). Therefore "zooming-in" at the corresponding location on the grid file will allow the user to visually compare both ENFs. It is clear that a modification has taken place between minutes 40 and 50 in the audio recording, as shown on the ENF, since its counterpart in the grid ENF does not have the same shape.

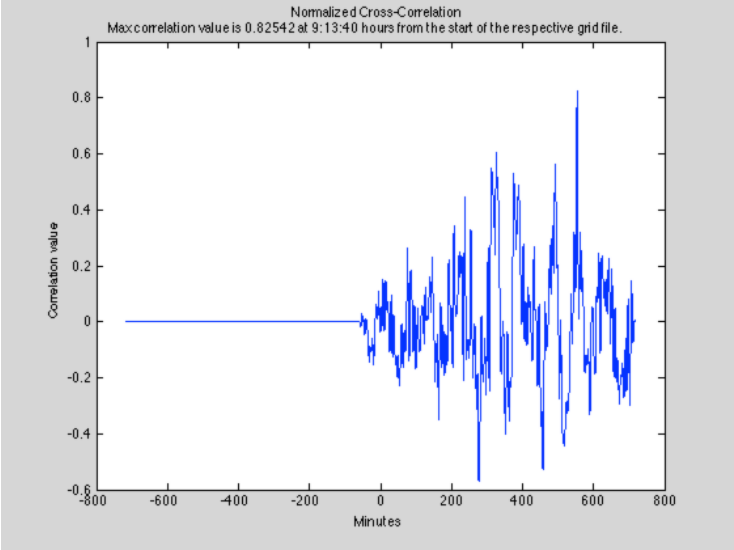

Figure 6.15 – Normalized correlation between modified/altered audio ENF and grid ENF

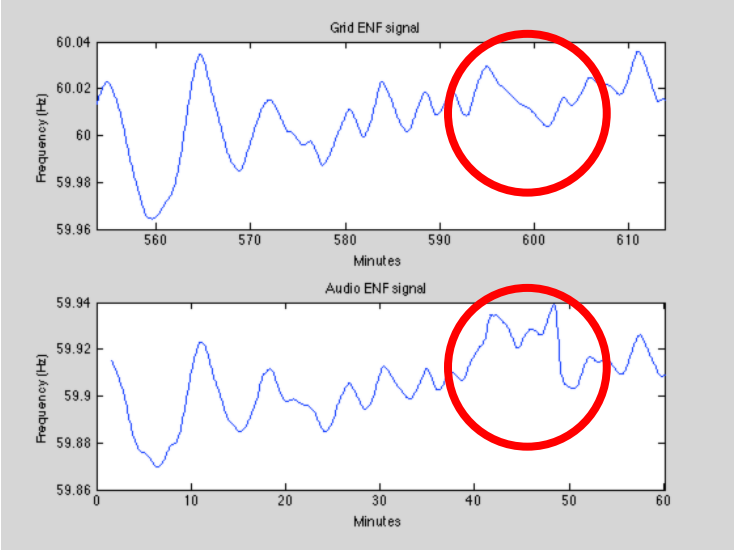

Figure 6.16 – a) Grid ENF at maximum correlation time, b) modified/altered audio ENF

Audio recording 60 minutes in length with 2 minutes of content modified, ENF extracted using STFT method around 60Hz peak

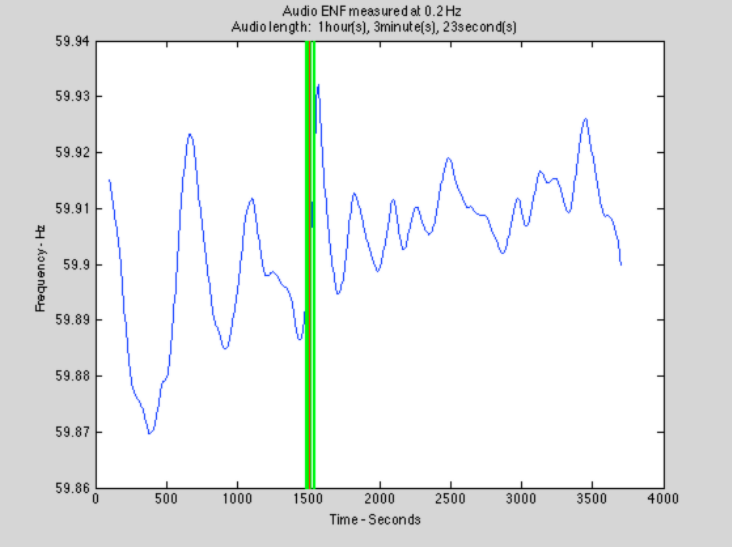

Figure 6.17 – Extracted ENF from audio signal modified for 2 minutes at minute 25.

Figure 6.18 – Warning sign indicating the time at which the alterations might have taken place, with most probable on being around 25 minutes

In this test, a 2 minutes modification took place in a 60 minutes audio recording, at 25 minutes from the start of the audio. The resulting ENF signal is shown in Figure 6.17 and was extracted using the STFT approach around the 60Hz peak. Since the alteration is quite short compared to the length of the audio recording (i.e. approximately 1 hour), it is expected that the normalized correlation value will be quite high but that the alteration time will nevertheless be detected and presented to the user via the warning pop-up sign, as shown in Figure 6.18, and via the green and red vertical lines.

Indeed, after computing the audio's ENF, the warning sign alerts the user with the most probable alteration is taking place at 25:10 minutes from the start of the recording. Simply looking at the audio's ENF cannot allow the user to provide a definite answer on the recording's authenticity. For this reason, the user will give his final verdict when

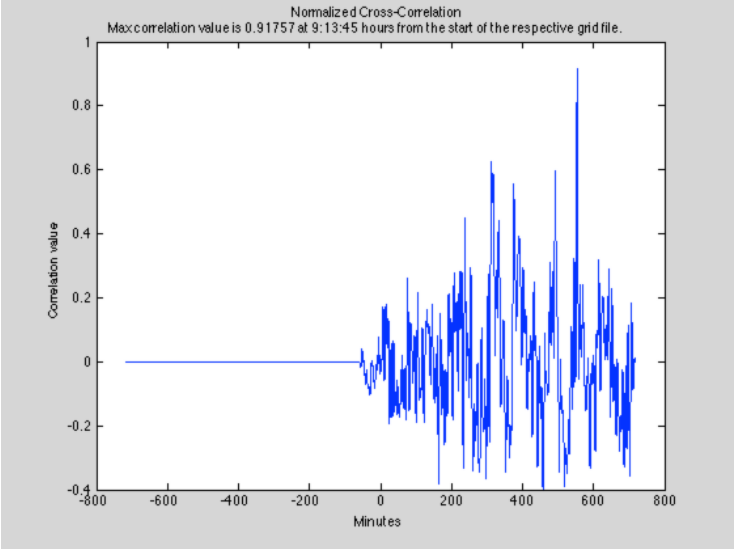

Figure 6.19 – Normalized correlation between modified audio ENF and grid ENF with correlation value of 0.917

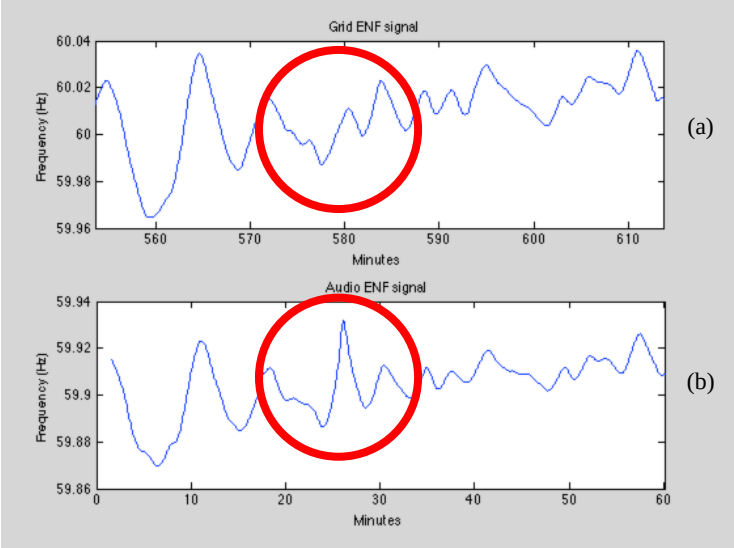

Figure 6.20 – a) Zoom on grid ENF at maximum correlation time of 9:13:45 hours, b) modified audio ENF

viewing the correlation results between both ENF signals, as shown in Figure 6.19. The correlation value obtained is high with 0.917 at 9:13:45 hours from the start of the grid (i.e. very similar time as the one computed for the previous examples). However, "zooming-in" on the grid ENF and analyzing both waveforms, see Figure 6.20, it is clear that the 2 minutes alteration has taken place in the audio recording, hence labelling the audio file as unauthentic, even though the correlation value is high.

Unaltered audio recording 60 minutes in length, audio ENF extracted using the AR parametric method around the 60Hz peak and grid ENF extracted using STFT method

This case scenario will use the AR approach described in Section 4.3 to show that the parametric method can in fact work for long signals as well as for shorter ones (to be covered in the upcoming examples). Figure 4.34 depicts the audio ENF obtained by AR

extraction of the information from the 60Hz peak. Figure 6.21 shows the normalized cross correlation graph and Figure 6.22 depicts the two ENF signals where the grid has been zoomed in at the corresponding correlation time of 9:13:40 hours from the start of the grid. The correlation value is 0.924 and is considered to be quite high, meaning that a good match is found somewhere within the grid ENF.

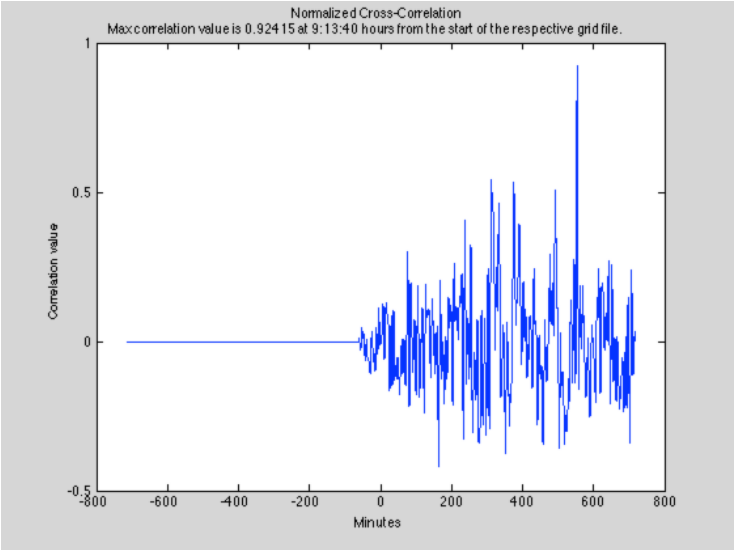

Figure 6.21 – Correlation result between audio ENF, using AR, and grid ENF using STFT

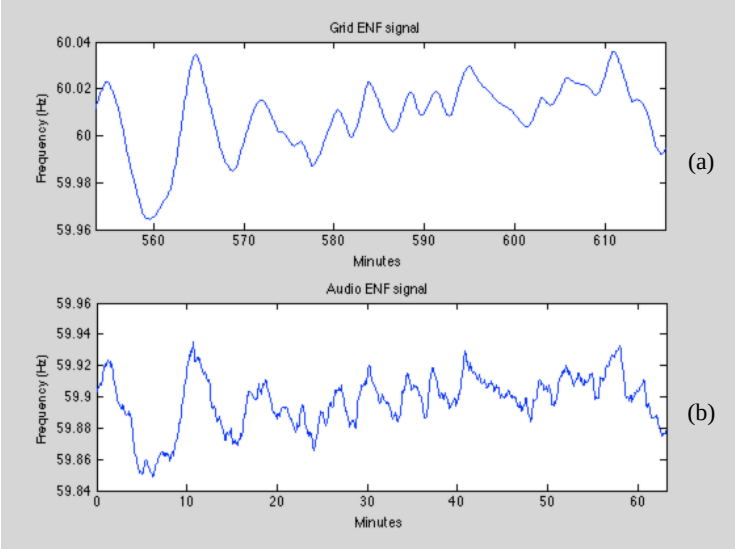

Figure 6.22 – a) Zoom at correlation time on the grid ENF, b) Audio ENF extracted using AR method

By visually analyzing both waveforms, it could be argued that the audio ENF obtained using the AR extraction method has more details than the STFT approach previously used (better resolution). Or alternatively, that it is noisier. In fact it is likely that both of these statements are true. It should be noted however that some trade-off is possible and one can obtain smoother ENF curves from AR method by simply changing the audio window size used in the algorithm, for example from 20

samples (4 samples at 0.2Hz) to 40 seconds, as described later in the thesis. The extracting methods previously developed have been purposely constructed using the same window shift, as to give the option to compare both methods with each other. However, each approach is different than the other in terms of resolution and it is recommended to correlate the grid ENF with the audio ENF that have been both extracted with the same method.

Unaltered audio recording 60 minutes in length, audio and grid ENF extracted using the AR parametric method around the 60Hz peak

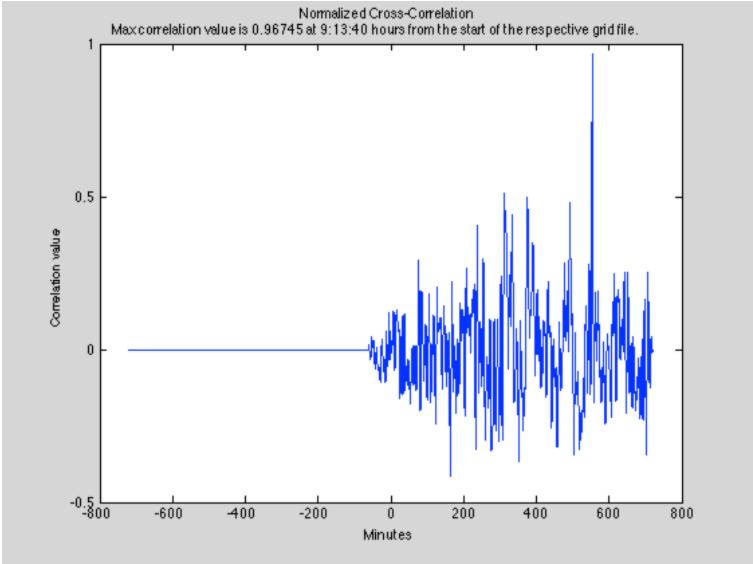

Figure 6.23 – Correlation computed for grid ENF and audio ENF that have been extracted with the AR method

For this example, both ENF signals have been computed using the AR approach, hence both having the same resolution. The correlation graph with maximum correlation value of 0.967 is plotted in Figure 6.23 and the ENF signals are depicted in Figure 6.24, where the grid ENF starts at the correlation time of 9:13:40 hours.

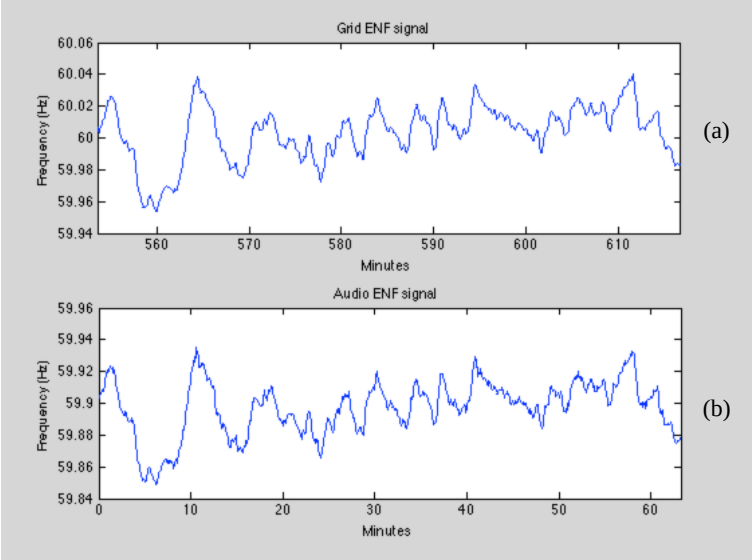

Figure 6.24 – a) Zoom at 9:13:40 hours on the grid ENF extracted using the AR method, b) audio ENF extracted using AR method

The correlation value has greatly increased compared to what was found when the ENF is extracted with the STFT method. It is now easier for the user to determine the recording's authenticity, which in this case shows that the audio file is definitely genuine. In the end, it is the user's choice to determine which method to use for the extraction of the grid ENF, each method having its own advantages and disadvantages.

Unaltered audio recording 15 minutes in length, extracted using AR method around 60Hz peak with varying the window size, grid ENF extracted using STFT method

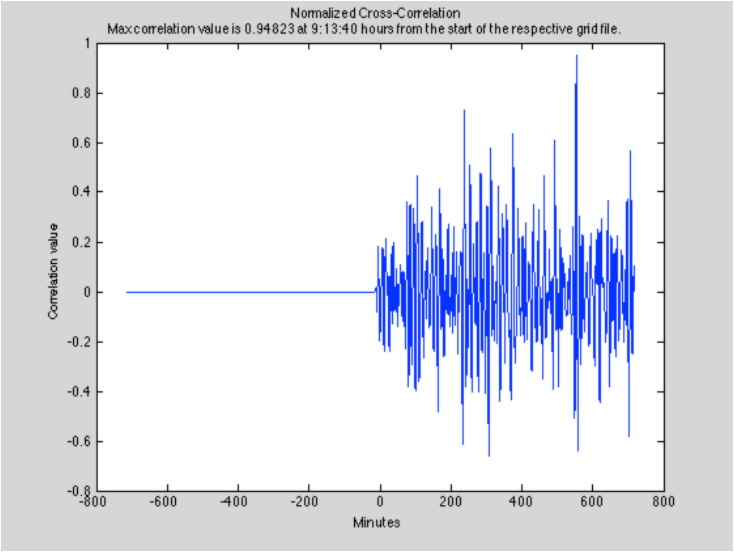

Figure 6.25 – Correlation result between audio ENF extracted with AR method using 20 seconds windows and grid ENF extracted with STFT method

In this example a recording of 15 minutes is analyzed with the AR method using two different window lengths. Extracting the ENF from 15 minutes of data (900 seconds) will provide 177 ENF samples when 20 seconds windows are used (one 20 sec. window +

176x5sec. shifts). The maximum correlation value obtained with the AR method using 20 seconds windows is 0.948 at 9:13:40 hours from the start of the grid, as shown in Figure 6.25. This is marginally better than the maximum correlation value that would be obtained if the STFT approach was utilized, giving a correlation of 0.945.

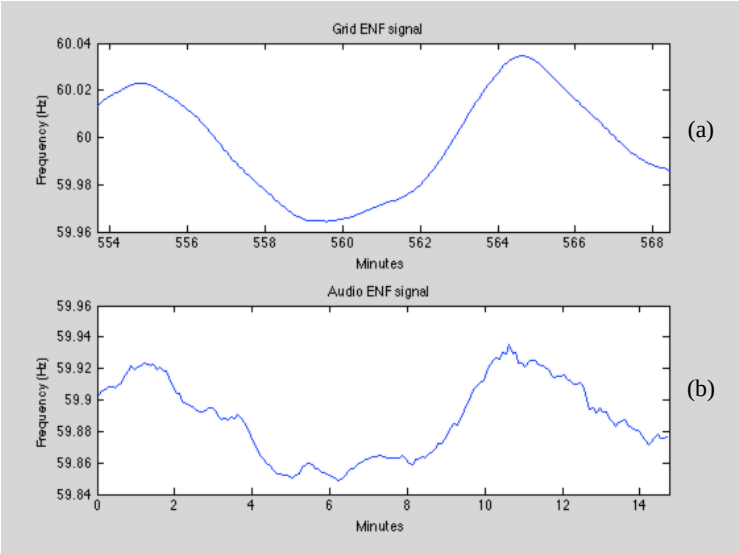

Figure 6.26 – a) Zoom on grid ENF at correlation time, b) audio ENF extracted using AR method with 20 seconds windows

As can be seen in Figure 6.26, the audio ENF's waveform is not as smooth as the ones obtained with the STFT method (for example the grid ENF in the same figure) since the algorithm was implemented using 20 seconds windows (i.e. meaning only 4 samples are analyzed for each window at 0.2Hz sampling rate) as opposed to 200 seconds windows for the STFT. As previously mentioned, the 20 seconds windows of the AR method lead to a better time resolution but also to noisier curves. It is possible to modify the window size to yield smoother ENF signals. For example, when increasing the window size to 40 seconds, the ENF signal obtained is significantly smoother, as depicted in Figure 6.28, and the correlation computed is further improved to 0.953 at 9:13:50, as can be seen in Figure 6.27. However, utilizing 40 seconds window would then require the audio recording file to have a minimum length of 40 seconds in order to provide 1 ENF sample, and one objective of the AR method is to reduce as much as possible the length of the digital recording needed for the extraction of the ENF signal.

The improvement provided by the AR method appears to be quite marginal for audio files of length 15 minutes or longer, with no alterations/modifications However, for shorter audio files such as 5 minutes in length, there can be significant improvements provided by the use of the AR method, as will be shown later. Unlike the STFT approach, the improved ENF resolution of the AR method can help to identify short alterations or modifications made in the audio signal.

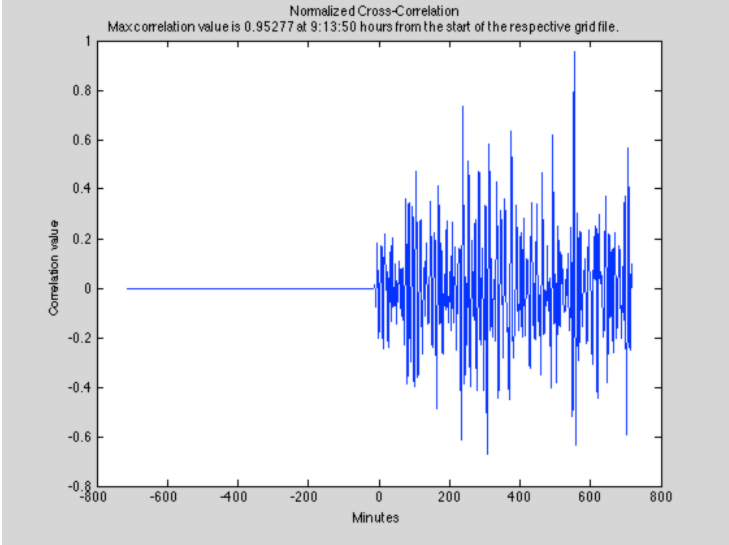

Figure 6.27 – Correlation result between audio ENF extracted with AR method using 40 seconds windows and grid ENF extracted with STFT method

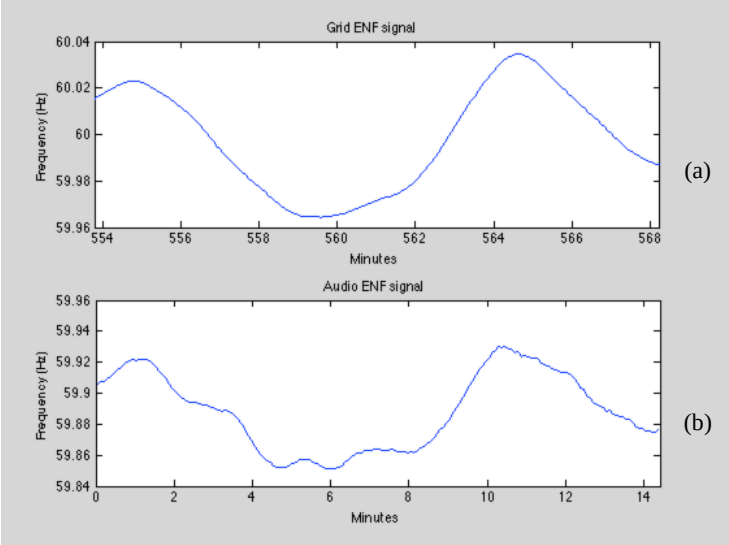

Figure 6.28 – a) Zoom on grid ENF at correlation time, b) audio ENF extracted using AT method with 40 seconds windows

Audio recording 15 minutes in length with 30 seconds of content modified, audio ENF extracted using AR, STFT and zero-crossing methods around 60Hz peak

In this example, 30 seconds were modified in an audio recording of length 15 minutes. The ENF extraction was first applied using the AR method with the default window length of 20 seconds. The correlation value obtained is 0.917 at 9:18:30 from the start of the grid, as shown in Figure 6.29, and a warning sign also appeared indicating that a modification likely took place at 5 minutes from the start of the audio file, refer to Figure 6.9. Indeed, the notification message will also warn the user of any probable alteration even when computing the correlation (due to the fact that the audio ENF can be loaded as a ".mat" file). When "zooming-in" on the grid ENF, as shown in Figure 6.30, at the corresponding time of 5 minutes the user would immediately notice the tampered part in the signal.

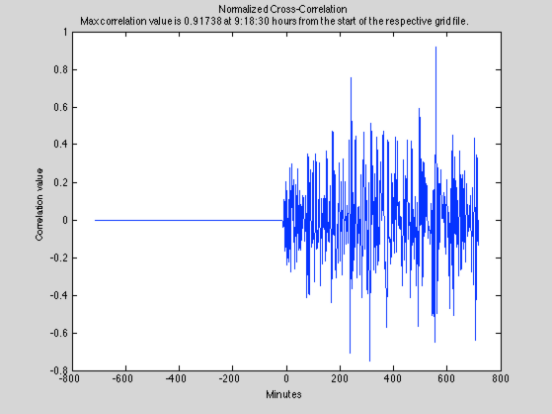

Figure 6.29 – Correlation result between audio ENF extracted using AR method and grid ENF

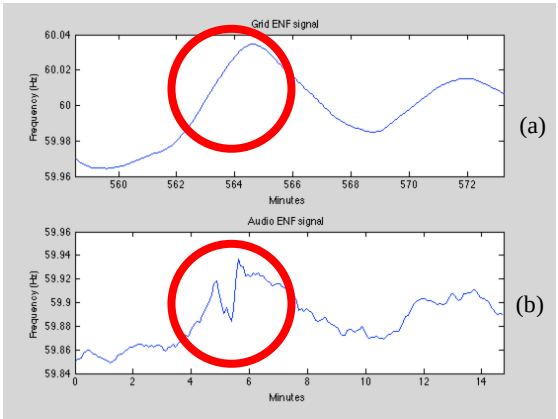

Figure 6.30 – Grid ENF at correlation time, b) modified audio ENF around 5 minutes

On the other hand, when the STFT method is applied for the audio's ENF extraction the algorithm does not output any warning message indicating that an alteration has occurred. Moreover the correlation computation provides the user with a correlation value of 0.851 at 4:1:50 hours from the start of the grid, as shown in Figure 6.31.

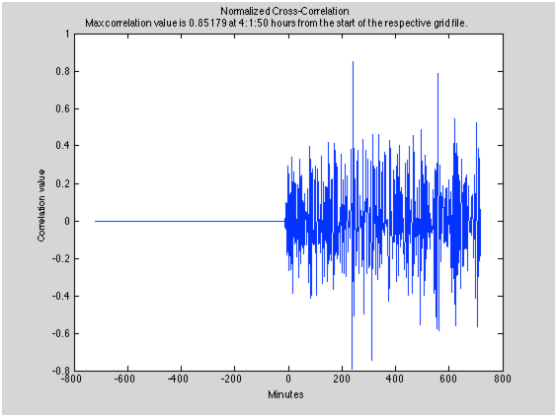

Figure 6.31 – Low correlation result between audio ENF extracted using STFT method and grid ENF and false correlation time obtained

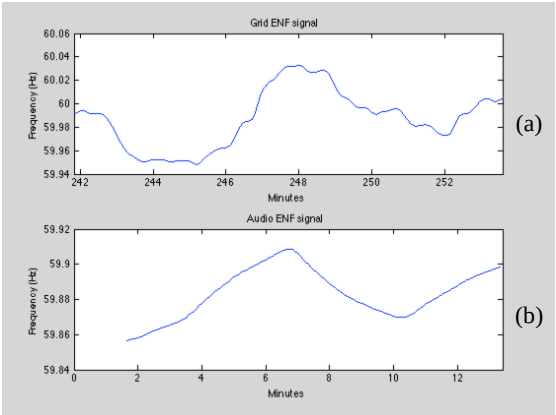

Figure 6.32 – a) Grid ENF at computed correlation time, b) altered audio ENF extracted with STFT method

The algorithm was unable to detect the 30 seconds tampering since the STFT method analyzes 200 seconds of data for every window, depicted in Figure 6.32 (poor time resolution). Therefore it must be said that the AR algorithm works better for the detection of short modifications made to a digital audio recording. The zero-crossing approach was found to be able to detect the alterations within the audio signal, since it also has a good time resolution. The corresponding results are shown in Figures 6.33 and 6.34, leading to a similar correlation value. Let us recall however that the complexity of the zero-crossing method is higher because of the sampling rate required.

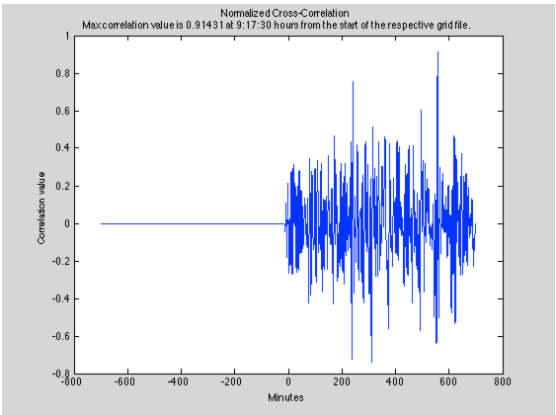

Figure 6.33 – Correlation result between audio ENF and grid ENF extracted with zero-crossing method

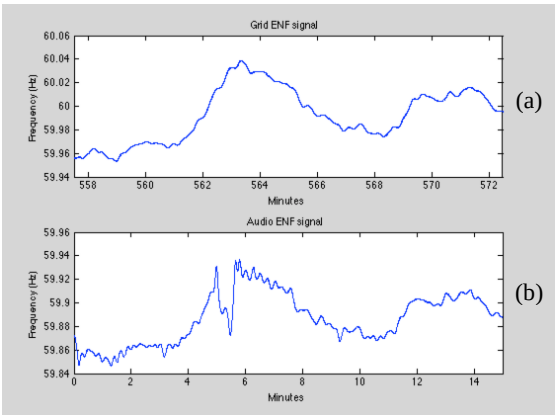

Figure 6.34 – a) Grid ENF at computed correlation time, b) altered audio ENF extracted with zero-crossing method

Comparing correlation values between a 1 hour grid ENF signal and the audio ENF extracted using the STFT, AR and zero-crossing method for a 5 minutes audio recording

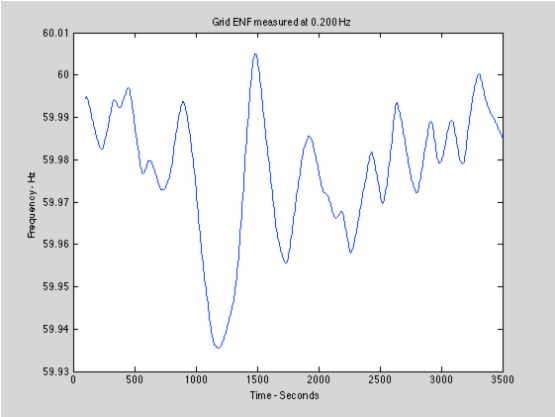

Figure 6.35 – Grid ENF extracted using STFT method for 1 hour of reference data

In this section a grid signal of length 1 hour, as shown in Figure 6.35, is used to compare the correlation values obtained when applying the STFT, AR and zero-crossing ENF extraction methods for a short audio signal of 5 minutes. Having a 12 hours reference grid ENF combined

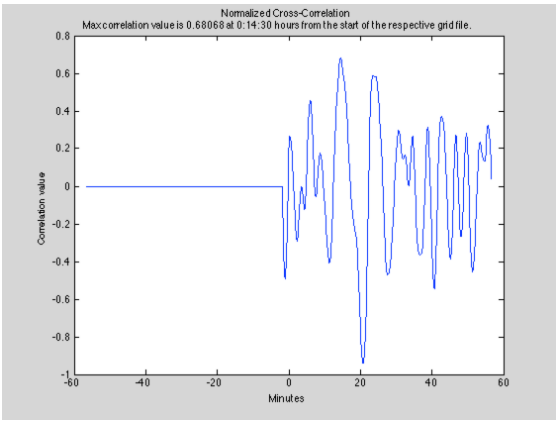

Figure 6.36 – Correlation result obtained using the STFT approach on a 5 minutes audio recording

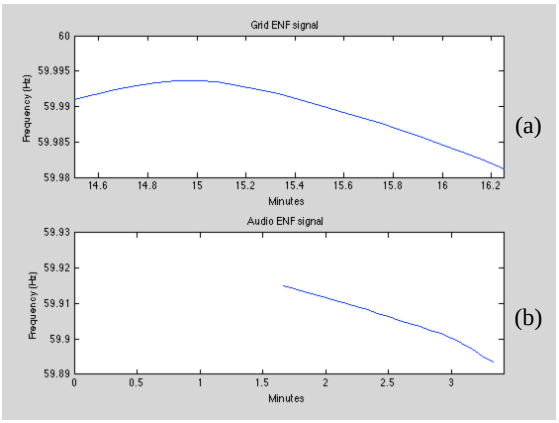

Figure 6.37 – a) Grid ENF at correlation time, b) audio ENF extracted using STFT method

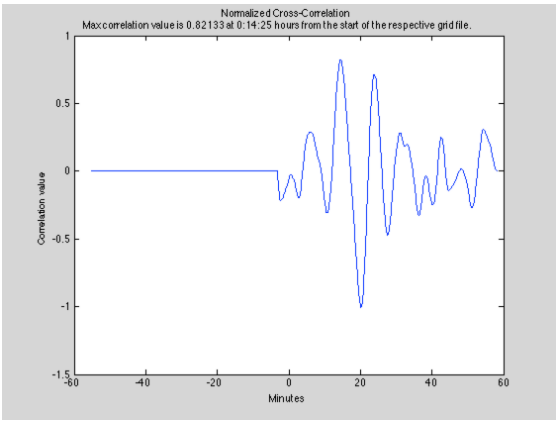

Figure 6.38 – Correlation result obtained using AR method on a 5 minutes audio recording

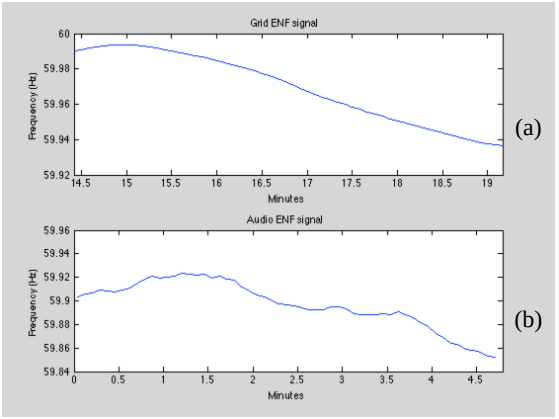

Figure 6.39 – a) Grid ENF at correlation time, b) audio ENF extracted using AR method

with a short audio signal ENF means that the chances of locating the correct correlation time are quite lower, because the odds that other segments in the 12 hours grid ENF will correlate with the short audio ENF are quite stronger. For this reason, the user can shorten the 12 hours original grid file into 1-hour segments, as applied here. The ENF extraction will produce more samples using the AR method with 20 seconds windows compared to the STFT method with 200 seconds windows (i.e. 57 points for the AR modeling with 1x20 seconds +56x5 seconds window shifts compared to 21 points with the STFT method with 1x200 + 20x5 seconds window shifts). The additional number of ENF samples for the AR method will lead to an improved correlation when comparing with the grid ENF. The correlation value obtained when employing the STFT approach is 0.689 at 14:30 minutes from the start of the grid, as shown in Figure 6.36, and the corresponding waveform is plotted in Figure 6.37.

As it can be seen, the ENF computed with the STFT method contains very few samples, the first one starting at 100 seconds (i.e. half of the 200 seconds window). On the other hand, the AR method gives an ENF signal with significantly more samples, as shown in Figure 6.39, hence leading to a better correlation value of 0.821 occuring at 14:25 minutes from the start start of the grid (a correlation value that is much more reliable than the 0.680 maximum correlation obtained using the STFT method).

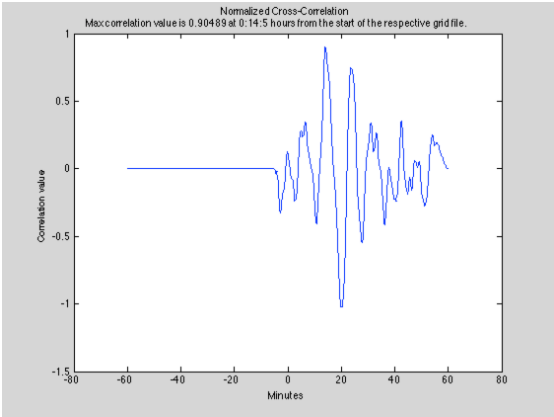

Figure 6.40 – Correlation result obtained using zero-crossing approach on 5 minutes audio recording

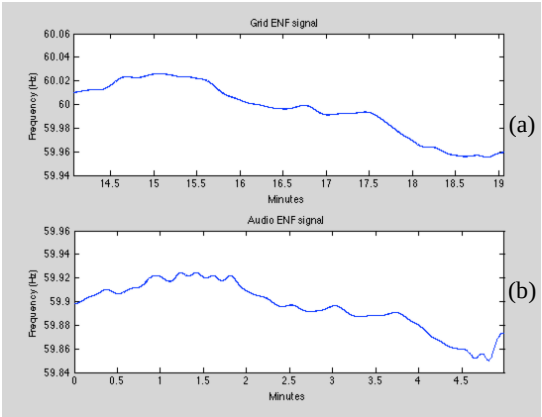

Figure 6.41 – a) Grid ENF at correlation time, b) audio ENF extracting using zero-crossing method

When applying the zero-crossing method, the correlation value obtained is actually the highest with 0.904 at 14:5 minutes from the start of the grid, as depicted in Figure 6.40. As Figures 6.34 and 6.41 show, the time resolution of the zero-crossing method appears to be even better than for the AR method (at the cost of higher computational complexity), although this increased time resolution also corresponds to more noisy curves.

Unaltered audio recording 15 minutes in length, ENF extracted using zero-crossing method and compared with a 1 hour grid ENF also extracted using zero-crossing

As previously said, comparing ENFs that are extracted using the zero-crossing

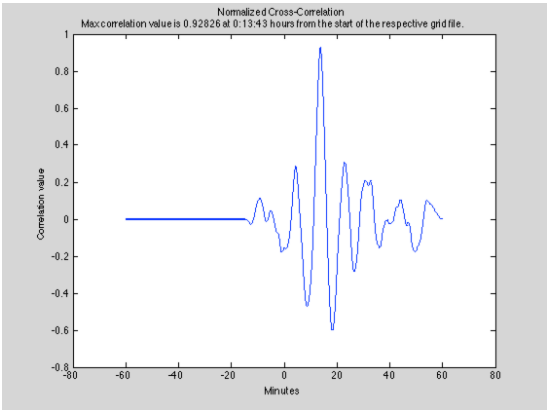

Figure 6.42 – Correlation result obtained using zero-crossing method on the grid signal and 15 minutes audio recording

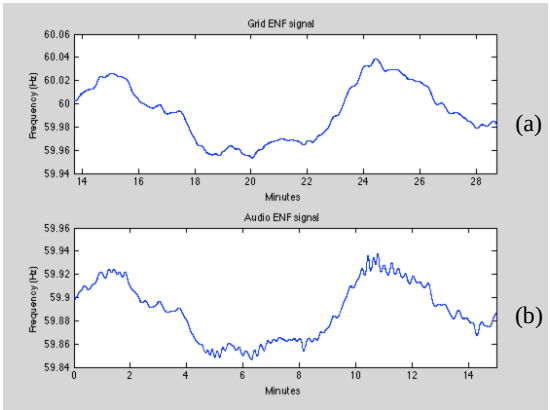

Figure 6.43 – a) Grid ENF, extracted using zero-crossing method, at correlation time, b) audio ENF extracted using zero-crossing method

method with a grid ENF obtained with the AR or STFT approach (or vice-versa) is not straightforward, due to different ENF sample rate (e.g. typically 0.2 Hz for STFT and AR, typically 120 Hz for zero-crossing). A comparison would require downsampling the ENF signal produced with the zero-crossing method, but severe downsampling on relatively short signals is not always applicable. In this example the zero-crossing approach is used for the ENF extraction of the 15 minutes unaltered audio recording and the hour-long grid file. The ENF waveforms are depicted in Figure 6.43. The correlation value obtained is 0.928 at 13:43 minutes from the start of the grid, as shown in Figure 6.42.

Applying the zero-crossing method can thus also provide decent results, as the audio ENF definitely has the same ENF shape as the grid ENF computed. The details are much more present in the grid ENF than what was previously computed with the STFT method, although it could also be argued that the method is more noisy. The complexity of the ENF extraction is high with the zero-crossing method.

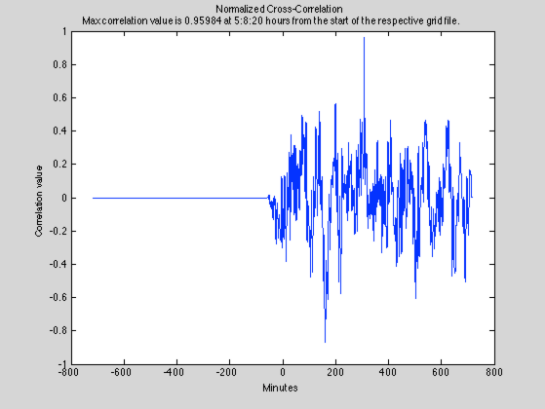

Figure 6.44 - Correlation result between grid ENF and audio ENF for the Marantz recorder

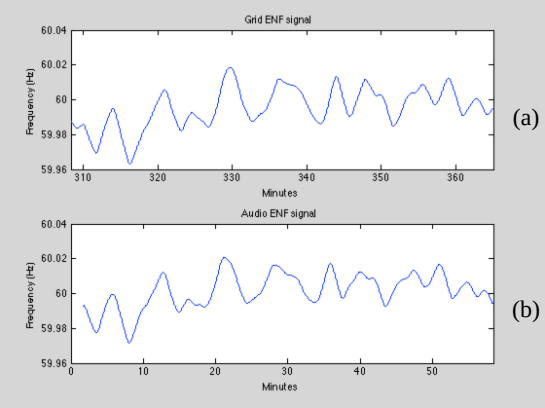

Figure 6.45 - a) Grid ENF at correlation time, b) audio ENF for the Marantz recorder

6.3.2. Marantz recorder (mains-powered mode)

Unaltered audio recording 60 minutes in length, ENF extraction using STFT method around the 60Hz peak

This section verifies the use of the ENF STFT extracting algorithm for the Marantz recorder in the mains-powered mode. The Marantz recorder samples the audio at 44.1kHz therefore the user must select the appropriate settings for the corresponding recording device from the GUI when computing the ENF. Performing the correlation between the audio ENF and the 12 hours grid ENF leads to a high normalized correlation result of 0.959 at a time of 5:8:20 hours from the start of the grid or at 1:17:21 GMT, as shown in Figure 6.44. The ENF waveform are depicted in Figure 6.45 and they prove that the ENF extraction algorithm works also for the Marantz recorder, even though the level of 60Hz ENF signal in this recorder is significantly lower than the Olympus WS210S or the Olympus DS150 models (as previously shown in Table 4.2).

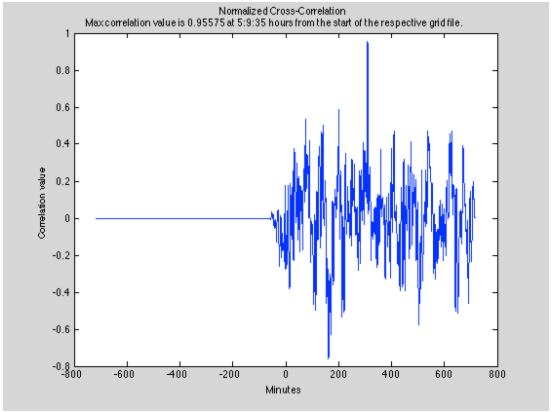

Figure 6.46 - Correlation result between grid ENF and audio ENF for the Olympus DS150 device

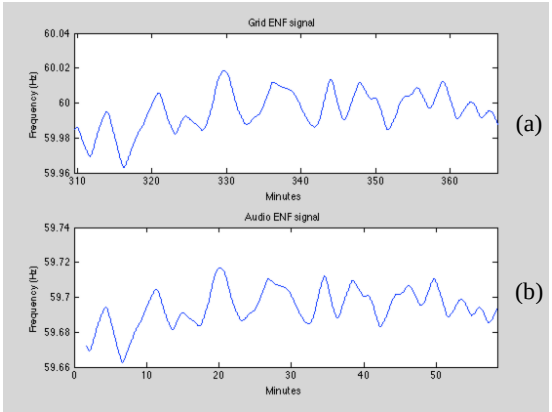

Figure 6.47 - a) Grid ENF at correlation time, b) audio ENF for Olympus DS150 device

6.3.3. Olympus DS150 audio recording device (mains-powered mode)

Unaltered audio recording 60 minutes in length, ENF extraction using STFT method around 60Hz peak

This section verifies the use of the ENF STFT extraction algorithm for the Olympus DS150 in the mains-powered mode. This device records the audio at a sampling rate of 8kHz. Computing the audio recording's ENF and then correlating it with the 12 hours grid will give a high-normalized correlation of 0.955 at 5:11:05 hours from the start of the grid or at 1:20:06 GMT (i.e. obtained using the time-stamp from the grid file), as shown in Figure 6.46. The ENF waveforms for both signals are depicted in Figure 6.47.

6.4. Compression analysis

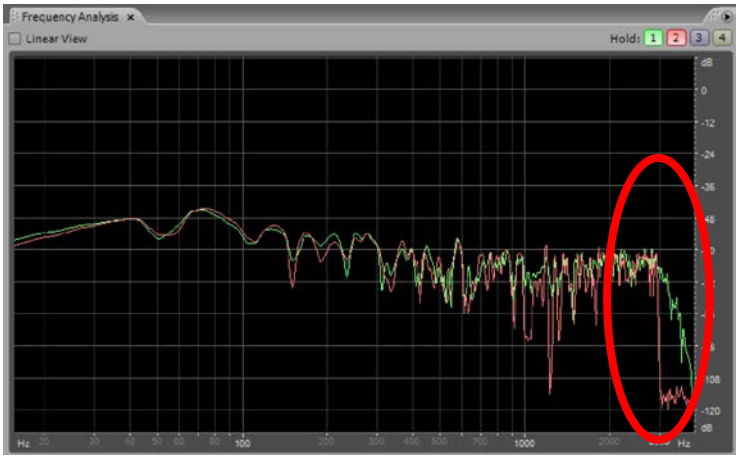

Figure 6.48 - Frequency analysis comparison between the ".wav" (green line) and ".mp3" (red line) format for the same recording

In this section, audio files that have been converted to ".mp3" format are being analyzed. Nowadays ".mp3" is widely used for compressing audio recordings. For this reason, the audio recording file of length 1 hour, used throughout this thesis, originally recorded at 8kHz and saved under a ".wav" format is converted to ".mp3" using the FFMPEG software. The program is a free software that runs using the command prompt in Windows. Once the ".wav" file is converted into ".mp3" at the lowest quality of 8kbps (i.e. compression eliminates most of

the unnecessary frequencies), both files are analyzed in Adobe Audition to verify where the frequencies have been modified, as shown in Figure 6.48.

It seems that the only frequencies affected and thus reduced are the ones above 3kHz. Therefore, the ENF extraction from the ".wav" file obtained by reconverting the ".mp3" should still work perfectly. Figure 6.49 depicts the ENF signal obtained from the original ".wav" file while Figure 6.50 depicts the ENF signal extracted from the ".mp3" file (re-converted to .wav file format for ENF extraction). Clearly both ENF signals extracted are nearly identical, meaning that the compression in ".mp3" has not greatly affected the ENF component located at 60Hz. It is also necessary to compute the correlation to see whether the correlation values are different from each other. The correlation graph plotted in Figure 6.51 corresponds to the original ".wav" file whose maximum correlation

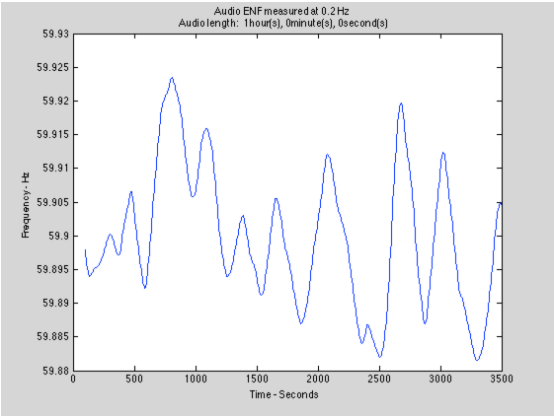

Figure 6.49 - Audio ENF extracted using STFT method from the original ".wav" file

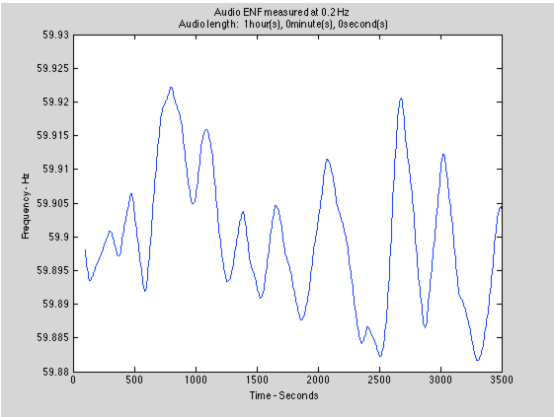

Figure 6.50 - Audio ENF extracted using the STFT method from the compressed ".mp3" file

value is 0.993 at 34:0 minutes from the start of the 2 hours grid signal. By "zooming-in" on the grid ENF at 34:0 minutes it is clear that the audio file has not been tampered with, thus labeling it authentic, refer to Figure 6.53. On the other hand, the correlation obtained from the compressed ".mp3" file results in a maximum correlation value of 0.991 also at 34:0 minutes from the start of the grid, as shown in Figure 6.52. The correlation value is

lower than what is obtained from the original ".wav" file, but it is, nonetheless, still clear to the user that the audio's ENF matches the grid's ENF, see Figure 6.54.

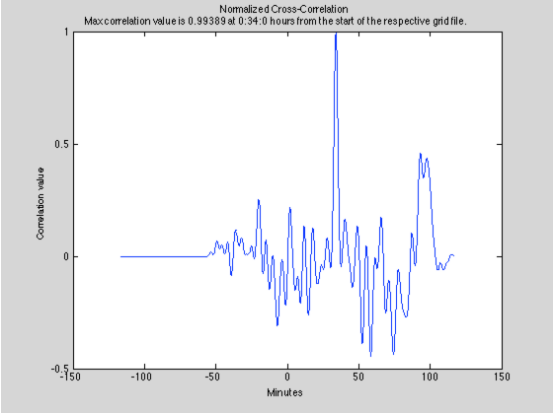

Figure 6.51 - Correlation result between 2 hours grid ENF and audio ENF from original ".wav" file

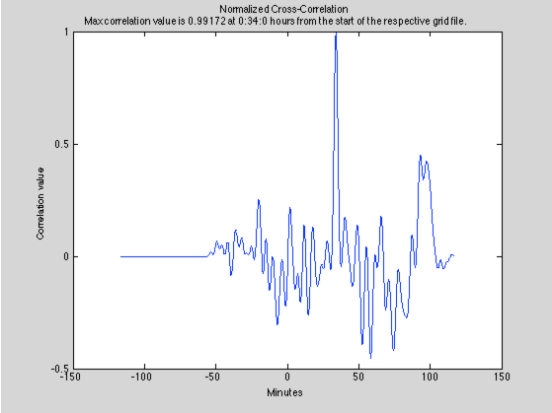

Figure 6.52 - Correlation result between grid ENF and audio ENF from compressed ".mp3" file

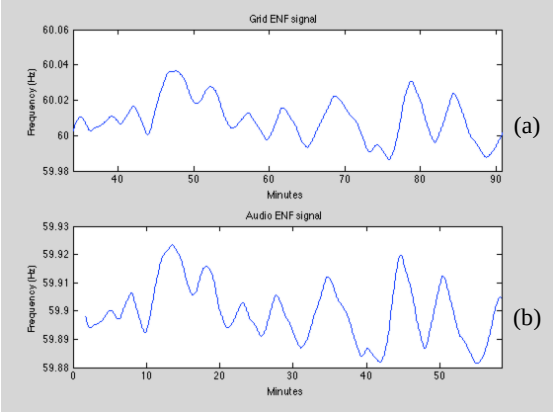

Figure 6.53 - a) Grid ENF at correlation time, b) audio ENF extracted from original ".wav" file

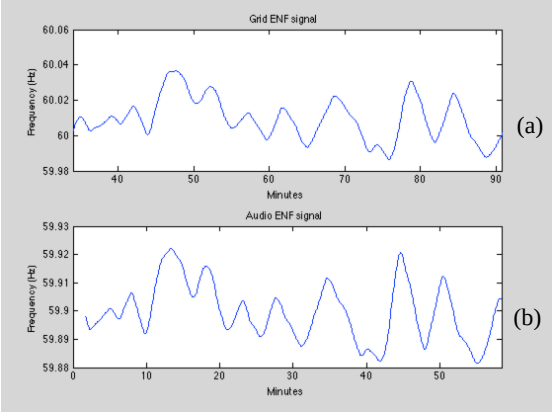

Figure 6.54 - a) Grid ENF at correlation time, b) audio ENF extracted from compressed ".wav" file

7. Conclusion

The objective of this thesis was to assess the feasibility and provide additional tools for the analysis of ENF signals, in order to determine whether an audio recording is authentic or not. Two methods that had previously been proposed for ENF extraction were implemented (STFT and zero-crossing), and we also proposed the novel use of an autoregressive (AR) parametric method for ENF extraction. We also developed a GUI in MATLAB for ENF extraction and matching, as well as a C-program and a probe to produce a grid ENF database. The STFT approach, based on an FFT analysis, is mostly useful for long signals (i.e. 15 minutes or longer, in order to have enough ENF samples to correlate with the grid ENF samples) as the use of long windows produces smooth and clean ENF curves. The proposed AR modelling method utilizing a parametric approach is intended for shorter recordings, although it is also applicable for longer recordings. The use of shorter windows provides great details in the ENF waveform and can therefore detect modified segments that are shorter in length, whereas the FFT is not capable. The zero-crossing method is also normally used for shorter signals, however this method can have a high computational load since it operates at higher sampling rates. Multiple case scenarios were tested and the ENF extraction proved to be robust for most of them. The problematic case of short alterations/modifications in long audio files, leading to strong overall correlation levels with the reference grid signal despite the audio alteration, can be handled by detecting the fast transitions in the audio file's ENF and showing a warning pop-up window to the user so he/she can verify the ENFs at the location of the detected alteration. The other problematic case of short audio files ENF (i.e. 10 minutes or less) to be compared with long grid files ENFs (12 hours), where multiple false

detections as “best match” could occur, can be handled by using the AR method (or the more expensive zero-crossing method) for ENF extraction and comparing the audio ENF with shorter segments from the grid signal (for example 1 hour segments instead of 12 hours).

7.1. Contributions

This thesis is based on an idea that was first proposed by Catalin Grigoras in 2002 [19]. The STFT and zero-crossing methods were implemented based on what was presented by the author. On the other hand the AR method developed in Section 4.3 is a new approach of tackling the extraction of the ENF fluctuations. Autoregressive models are based on parametric models that extract the statistical model of the corresponding signal yielding to the estimation of the output’s PSD. The proposed method was found to be effective for dealing with short signals or short alterations, with a reasonably low complexity (compared to the zero-crossing method). Thus it is a novel contribution.

The C-program developed is also a contribution to the thesis and it helped to capture the grid ENF from the probe. The gathered data is converted into a “.raw” format file containing 12 hours of ENF information. In addition, the C-program also creates a time-stamp that is also saved along the “.raw” grid file and is intended to be used by the GUI for time and date confirmation. The building of 3 probes is also a minor contribution from the work of this thesis.

The interface built in MATLAB is a very important tool for the extraction of the ENF signal. The GUI is simple to use yet very efficient. The user can select multiple extraction methods as well as the peak to be analyzed (i.e. 60Hz, 120Hz or 180Hz). Once the ENF is computed, the user has the option of saving the ENF into a “.mat” file, hence

eliminating the need to re-run the algorithm later on. The correlation algorithm is executed on the click of a button and the results are clearly presented on the computer's screen, indicating the correlation value, the time/date of the corresponding maximum correlation value (location of the best match in the grid ENF). If the time-stamp file is found then the date/time are with respect to the GMT time reference, otherwise the result is with respect to the start of the grid ENF file. If a major frequency difference is found between two successive ENF samples, the algorithm will warn the user (on a popping message) with all the probable alterations that might have taken place throughout the audio file and will indicate the locations with vertical lines (red line indicates the major alteration and the green lines indicate the rest of the alterations) on the plotted ENF waveform.

7.2. Recommendation and future work

After completing and achieving the objectives of the thesis and analyzing the presented results, the following recommendations could bring improvements to the extraction of the ENF signals from digital recording:

- The sampling rate utilized in the C-program for the sampling of the captured grid ENF could be reduced (meeting the user's preference), if a proper sound board is available
- A double buffer could be included in the C-program, along the 10 seconds capturing rate, which will reduce even further the chance of missing any samples during the recording process
- Rather than only using a correlation based matching method, to develop a second method based on Minimum Squared Error (MSE) that will detect a

match between the audio ENF and the grid ENF, perhaps making the matching even more robust as both algorithms can work in parallel.

- Eventually include in the GUI an option that allows the user to utilize audio recordings that are not in ".wav" format, which the algorithm will convert when the program is run.
- Determine the threshold value for detecting a probable alteration by obtaining enough information from a large number of tests, hence providing a statistical distribution of data for the ideal value

8. References

[1] C. Grigoras, J. M. Smith, and C. W. Jenkins, "Advances in ENF Database Configuration for Forensic Authentication of Digital Media," *Audio Engineering Society Conference*, 2011.

[2] S. Merlo, "Establishment of the individual characteristics of magnetic recording systems for identification purposes," *Problems in Forensic Sciences*, vol. 47, pp. 20-39, 2001.

[3] Oxford English Dictionary. (2013). *Authentic*. Available: http://www.oed.com/view/Entry/13314

[4] Audio Engineering Society, "AES recommended practice for forensic purposes — Managing recorded audio materials intended for examination," AES Standard AES27-1996, 2007.

[5] Oxford English Dictionary. (2013). *Forum*. Available: http://www.oed.com/view/Entry/73767

[6] S. Bell, *Crime and Circumstance: Investigating the History of Forensic Science*. London: Praeger, 2008.

[7] H. Buker, *Fraudulent Forensic Evidence: Malpractice in Crime Laboratories*: LFB Scholarly Publishing LLC, 2012.

[8] R. Morelle. *The hum that helps to fight crime*. 2012. Available: http://www.bbc.co.uk/news/science-environment-20629671

[9] B. E. Koenig, "Authentication of Forensic Audio Recordings," *Journal of the Audio Engineering Society*, vol. 38, No. 1/2, pp.3-33, 1990.

[10] B. E. Koenig and D. S. Lacey, "Forensic Authentication of Digital Audio Recordings," *Journal of the Audio Engineering Society*, vol. 57, No. 9, pp.662-695, 2009.

[11] Y. Liu, Z. Yuan, P. N. Markham, R. W. Conners, and Y. Liu, "Application of Power System Frequency for Digital Audio Authentication," *IEEE Transactions on Power Delivery*, vol. 27, No. 4, pp. 1820-1828, 2012.

[12] C. Grigoras, "Digital Audio Recording Analysis: The Electric Network Frequency (ENF) Criterion," *International Journal of Speech Language and the Law*, vol. 12, No. 1, pp. 64-76, 2005.

[13] O. Ojowu, J. Karlsson, J. Li, and Y. Liu, "ENF Extraction From Digital Recordings Using Adaptive Techniques and Frequency Tracking," *IEEE Transactions on Information Forensics and Security*, vol. 7, No. 4, pp. 1330-1338, 2012.

[14] B. Gerazov, Z. Kokolanski, G. Arsoc, and V. Dimcev, "Tracking of electrical network frequency for the purpose of forensic audio authentication," *13th international conference on Optimization of Electrical and Electronic Equipment*, pp. 1164-1169, 2012.

[15] F.-C. Chang and H.-C. Huang, "A Study on ENF Discontinuity Detection Techniques", *Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 9-12, 2011.

[16] R. W. Sanders, "Digital Audio Authenticity Using the Electric Network Frequency," *Proceedings of Audio Engineering Society Conference*, vol. 57, No. 9, 2008.

[17] C. Grigoras, "Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis," *Forensic Science International,* vol. 167, pp. 136-145, 2007.
[18] E. B. Brixen, "Techniques for the authentication of digital audio recordings," *Proceedings Audio Engineering Society Conference,* vol. 122, 2007.
[19] C. Grigoras, "The Electrical Network Frequency Analysis Over Audio Recording," *Proceedings of Advanced Topics in Electrical Engineering Conference (ATEE),* 2002.
[20] C. Grigoras, "Forensic Analysis of Digital Audio Recordings - The Electric Network Frequency Criterion," *Forensic Science International,* vol. 136, pp.368-369, 2003.
[21] C. Grigoras and J. M. Smith, "Advances in ENF Analysis for Digital Media Authentication," *Audio Engineering Society 46th International Conference,* pp. 108-115, 2012.
[22] M. Kajustra, A. Trawinska, and J. Hebenstreit, "Application of the Electrical Network Frequency (ENF) Criterion - A case of a digital recording," *Forensic Science International,* vol. 155, pp. 165-171, 2005.
[23] R. C. Maher, "Audio Forensic Examination: Authenticity, enhancement, and interpretation," *IEEE Signal Processing Magazine* vol. 26, pp. 84-94, 2009.
[24] C. Grigoras, "Applications of ENF Analysis in Forensic Authentication of Digital Audio and Video Recordings," *Journal of the Audio Engineering Society,* vol. 57, No. 9, 2009.
[25] North American Electric Reliability Corporation. *About NERC: Understanding the Grid* (2013). Available: http://www.nerc.com/page.php?cid=1%7C15
[26] North American Electric Reliability Corporation. (2012). *Key Players: Regional Entities*. Available: http://www.nerc.com/AboutNERC/keyplayers/Documents/NERC_Interconnections_Color_072512.jpg
[27] North American Electric Reliability Corporation, "Balancing and Frequency Control: A Technical Document Prepared by the NERC Resources Subcommittee," Princeton, NJ, 2011.
[28] North American Electric Reliability Corporation, "Control Area Concepts and Obligations," *NERC Reliability Criteria for Interconnected Systems Operation,* 1992.
[29] North American Electric Reliability Corporation. *Compliance: About Compliance* (2012). Available: http://www.nerc.com/page.php?cid=3%7C249
[30] M. Amin and J. Stringer, "The Electric Power Grid: Today and Tomorrow," *MRS Bulletin,* vol. 33, pp. 399-407, 2008.
[31] North American Energy Standards Board, "Manual Time Error Correction," *NAESB WEQ Manual Time Error Correction Standards,* 2005.
[32] C. Jenkins, J. Leroi, and J. Steinhour, "Advances in Electric Network Frequency Acquisition Systems and Stand Alone Probe Applications for the Authentication of Digital Media," *Audio Engineering Society 46th International Conference,* pp. 89-102, 2012.
[33] M. L. Hiles, R. G. Olsen, K. C. Holte, D. R. Jensen, and K. L. Griffing, "Power Frequency Magnetic Field Management Using a Combination of Active and

Passive Shielding Technology," *IEEE Transactions on Power Delivery*, vol. 13, No. 1, pp.171-179, 1998.

[34] W. E. Feero, "Magnetic Field Management," *Proceedings of the Scientific Workshop on the Health Effect of Electric and Magnetic Fields on Workers,* pp. 167-184, 1991.

[35] Wisconsin Public Service, "Gas and Electric Service Manual for Wisconsin and Michigan - EMF Guidelines," pp. 6-6, 2010.

[36] W&W Radiological and Environmental Consultant Services Inc, "Guidance Manual for the Preparation of an EMF Management Plan for the City of Toronto," *EMF Management Plan Guidance Manual*.

[37] *Understanding FFT Windows - University of Wurzburg* (2003). Available:http://www.physik.uni-wuerzburg.de/~praktiku/Anleitung/Fremde/ANO14.pdf

[38] Encyclopedia.com. *Walker, Gilbert Thomas - Complete Dictionary of Scientific Biography 2008*. Available: http://www.encyclopedia.com/doc/1G2-2830906190.html

[39] S. S. Haykin, *Adaptive Filter Theory - Fourth Edition*: Prentice Hall, 2002.

[40] A. J. Cooper, "The Electric Network Frequency (ENF) as an aid to Authenticating Forensic Digital Audio Recordings - an Automated Approach," *Proceedings Audio Engineering Society 33rd International Conference*, 2008.

[41] A. J. Cooper, "An Automated Approach to the Electric Network Frequency (ENF) Criterion: Theory and Practice," *International Journal of Speech Language and the Law*, vol. 16, No. 2, pp.193-218, 2009.

[42] A. J. Cooper, "Further Considerations for the Analysis of ENF Data for Forensic Audio and Video Applications," *International Journal of Speech Language and the Law*, vol. 18, No. 1, pp.99-120, 2011.

[43] C. Grigoras, D. Rappaport, and J. M. Smith, "Analytical Framework for Digital Audio Authentication," *Audio Engineering Society 46th International Conference*, pp. 123-133, 2012.

[44] A. J. Cooper, "Detectiong Butt-Spliced Edits in Forinsic Digital Audio Recordings," *39th International Conference: Audio Forensics: Practices and Challenges*, 2010.