MULTIMEDIA AUTHENTICATION USING ELECTRIC NETWORK FREQUENCY

Georgios D. Karantaidis
Dipl. Eng., M.Sc. in Geoinformatics

Supervisor: Prof. Dr. Eng. Constantine Kotropoulos

Department of Informatics
Faculty of Sciences
Aristotle University of Thessaloniki

This dissertation is submitted in partial fulfillment for the degree of
Master of Science in Informatics

Thessaloniki 2017

Acknowledgements

I wish to thank my advisor Prof. Dr. Eng. Constantine Kotropoulos, who has been a constant source of encouragement to me during the preparation of this thesis. He has always been ready to give generously of his time and I am deeply grateful for all the advice which I have received from him.

Abstract

Electric Network Frequency (ENF) analysis provides useful forensic evidence for multimedia authentication, such as time determination or verification of audio and video recordings, authenticity inspection, and forgery detection. It is based on the extraction of ENF fluctuations, which occur around the nominal value in a random, non-periodic way, due to the differences between demanded and produced power load. Frequency fluctuations are compared to existing ground truth data obtained from the power grid during the same period. In this thesis, a systematic study of refined periodogram, refined filter-bank, and high-resolution spectral estimation methods is conducted for ENF extraction. A comparison is carried out between these spectral estimation techniques and the traditional method of Short-Time Fourier Transform, as well as other methods proposed in the literature, such as Maximum-Likelihood estimation, spectral estimation based on dynamic programming, etc. All parameters used through the entire process and the way they influence ENF estimation are thoroughly studied. A systematic study is also carried out in order to measure how each algorithm and raw signal filtering affect the computational time. The experiments demonstrate that the developed spectral estimation techniques provide accurate ENF extraction, achieving a matching accuracy, measured by the correlation-coefficient between the extracted ENF and the ground truth, which exceeds 99%.

Contents

List of Figures

List of Tables

Dedicated to my family.

Εκτενής περίληψη

Η παρούσα μεταπτυχιακή διπλωματική εργασία εκπονήθηκε στα πλαίσια του Προγράμματος Μεταπτυχιακών Σπουδών «Πληροφορική και Επικοινωνίες» και ειδικότερα της κατεύθυνσης «Ψηφιακά Μέσα και Υπολογιστική Νοημοσύνη» του Τμήματος Πληροφορικής, του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης από τον Διπλωματούχο Αγρονόμο και Τοπογράφο Μηχανικό Γεώργιο Καρανταΐδη, υπό την επίβλεψη του Καθηγητή Κωνσταντίνου Κοτρόπουλου. Στόχος της παρούσας μεταπτυχιακής διπλωματικής εργασίας είναι η εξέταση, ανάλυση και αξιολόγηση του ελέγχου αυθεντικότητας πολυμέσων με τη χρήση της Συχνότητας Ηλεκτρικού Δικτύου (ΣΗΔ)($ElectricNetworkFrequency - ENF$). Η ΣΗΔ στις Ηνωμένες Πολιτείες είναι $60\,Hz$, ενώ στην Ευρώπη είναι $50\,Hz$. Οι τιμές της συχνότητας αυτής δεν είναι σταθερές, αλλά παρουσιάζουν μικρές διακυμάνσεις κατά τη διάρκεια του χρόνου. Οι μικρές αυτές αποκλίσεις οφείλονται στη στιγμιαία διαφορά μεταξύ της προσφοράς από τους παρόχους και της ζήτησης από τους καταναλωτές ηλεκτρικής ισχύος. Οι διακυμάνσεις της ΣΗΔ δεν παρουσιάζουν περιοδικότητα και συνεπώς δεν είναι δυνατό να προβλεφθούν. Το γεγονός αυτό σε συνδυασμό με την ιδιότητά τους να παραμένουν σταθερές μέσα στο ίδιο ηλεκτρικό δίκτυο, καθιστούν τη ΣΗΔ ένα αποτελεσματικό μέσο ελέγχου αυθεντικότητας πολυμέσων. Πολλές έρευνες επικεντρώνονται αφενός στην αποτελεσματικότερη εξαγωγή της ΣΗΔ από τις καταγραφές στις οποίες είναι ενσωματωμένη και αφετέρου, στη διαδικασία του ελέγχου αυθεντικότητας με την αντιπαραβολή της εξαχθείσας ΣΗΔ και της ΣΗΔ αναφοράς (γρουνδ τρυτη). Η εξαγωγή της ΣΗΔ είναι μία διαδικασία που απαιτεί μεγάλη ακρίβεια και αποτελεσματικότητα, διότι η συχνότητα πολλές φορές καθίσταται δύσκολο να ανιχνευθεί εξαιτίας του θορύβου, ειδικά στα σήματα ομιλίας. Η προαναφερθείσα ΣΗΔ αναφοράς μετριέται με μεγάλη ακρίβεια από τους παρόχους και

διαχειριστές ηλεκτρικών δικτύων. Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από το Πανεπιστήμιο της Φλόριντα και είναι δύο ειδών. Το πρώτο σήμα προέρχεται από την απευθείας καταγραφή της ΣΗΔ μέσω της κάρτας ήχου φορητού υπολογιστή που συνδέεται απευθείας με την πρίζα. Στο σήμα αυτό η παρουσία της ΣΗΔ είναι έντονη. Το δεύτερο σήμα είναι μια καταγραφή ομιλίας από το μικρόφωνο του φορητού υπολογιστή. Στο σήμα αυτό η ΣΗΔ είναι ασθενής και μπορεί να εντοπισθεί μόνο στη δεύτερη αρμονική της.

Στην εργασία παρουσιάζονται και αναλύονται εκτενώς: 1) η διαδικασία της προεπεξεργασίας των δεδομένων, η οποία παίζει καταλυτικό ρόλο στην αποτελεσματική εξαγωγή του σήματος της ΣΗΔ και 2) η εφαρμογή μεθόδων φασματικής ανάλυσης για την εξαγωγή της ΣΗΔ και για τον έλεγχο αυθεντικότητας της υπό εξέταση καταγραφής. Ο έλεγχος αυθεντικότητας σχετίζεται με το εάν έχει αλλοιωθεί το πρωτογενές πολυμεσικό σήμα, λ.χ. με προσθήκη πληροφορίας που καταγράφηκε σε διαφορετικές χρονικές στιγμές, όπως τεκμαίρεται από την ανάλυση του ίχνους της ΣΗΔ, αλλά και σε διαφορετικές περιοχές συσχετίζοντας τις τιμές της ΣΗΔ με τις καταγραφές αναφοράς σε διαφορετικά σημεία του ηλεκτρικού δικτύου. Επιπρόσθετα, παρουσιάζονται λεπτομερώς οι χρόνοι που χρειάστηκαν για την εξαγωγή της ΣΗΔ, για τα δύο είδη σημάτων που χρησιμοποιήθηκαν, της αρμονικής περί της οποίας φιλτραρίστηκε με ζωνοπερατό φίλτρο το υπό μελέτη σήμα, καθώς και του μεγέθους των επικαλυπτόμενων παραθύρων που χρησιμοποιήθηκαν. Η διαδικασία της προεπεξεργασίας των πρωτογενών δεδομένων ξεκινά με την υποδειγματολειψία του αρχικού δείγματος στα $441\,Hz$, ώστε να μειωθεί η υπολογιστική πολυπλοκότητα των διαδικασιών, κρατώντας παράλληλα την απαραίτητη πληροφορία για τις αρμονικές της ΣΗΔ. Στη συνέχεια, εφαρμόζεται ένα ζωνοπερατό φίλτρο, ώστε να παραμείνουν μόνο οι συχνότητες που είναι απαραίτητες γύρω από μια αρμονική, π.χ. την πρώτη αρμονική στα $60\,Hz$. Πιο συγκεκριμένα οι συχνότητες αποκοπής του ζωνοπερατού φίλτρου ορίζονται στα $59.95 - 60.05\,Hz$ για την πρώτη αρμονική, $119.95 - 120.05\,Hz$ για τη δεύτερη και $179.95 - 180.05\,Hz$ για την τρίτη αρμονική. Στην βιβλιογραφία οι περισσότερες εργασίες περιορίζονται στην ανάλυση φιλτραρισμένων σημάτων περί την πρώτη αρμονική, γεγονός που οδηγεί σε εσφαλμένες εκτιμήσεις τις ΣΗΔ στις περιπτώσεις σημάτων ομιλίας, όπου παρατηρούνται έντονες παρεμβολές και θόρυβος. Οι προαναφερθείσες συ-

χνότητες αποκοπής περί την τρίτη αρμονική οδήγησαν σε καλύτερη εξαγωγή της ΣΗΔ. Ένας ακόμη παράγοντας που επηρεάζει σημαντικά την ποιότητα του εξαγόμενου σήματος κατά τη διάρκεια της προεπεξεργασίας είναι η τάξη του ζωνοδιαβατού φίλτρου. Ειδικότερα για την περίπτωση του σήματος ομιλίας η τάξη του φίλτρου διαδραματίζει σημαντικό ρόλο, καθώς μία λανθασμένη επιλογή μπορεί να οδηγήσει σε κακή εξαγωγή της ΣΗΔ και κατά συνέπεια σε λανθασμένη εκτίμηση της αυθεντικότητας των δεδομένων. Για τη σχεδίαση του ζωνοδιαβατού φίλτρου επιλέχθηκε ένα φίλτρο πεπερασμένης κρουστικής απόκρισης και χρησιμοποιήθηκε το παράθυρο *Hamming*. Στη συνέχεια επιλέγεται ο αριθμός των επικαλυπτόμενων τεμαχίων στα οποία χωρίζεται το σήμα πολλαπλασιάζοντας με ένα ορθογώνιο παράθυρο. Η διάρκεια του παραθύρου που χρησιμοποιήθηκε είναι για το πρώτο σήμα $20\,sec$ ή $40\,sec$ και για το δεύτερο σήμα $33\,sec$ ή $50\,sec$. Κατόπιν, για κάθε τεμάχιο υπολογίζεται η πυκνότητα φάσματος ισχύος με μία πληθώρα μεθόδων, όπως: ο Βραχυχρόνιος Μετασχηματισμός *Fourier* ($Short-Time\,Fourier\,Transform$), η Εκτίμηση Παραμέτρων Σήματος με Αμετάβλητες ως προς την Περιστροφή Τεχνικές ($Estimation\,of\,Signal\,Parameters\,via\,Rotational\,Invariant\,Techniques\,-\,ESPRIT$), η μέθοδος *Capon*, η Ταξινόμηση Πολλαπλών Σημάτων ($Multiple\,Signal\,Classification-MUSIC$), και η Μέθοδος Μέγιστης Πιθανοφάνειας ($Maximum-Likelihood\,Estimation-MLE$). Στην παρούσα εργασία μελετώνται, επίσης, οι μέθοδοι τροποποιημένου περιοδογράμματος, όπως οι μέθοδοι *Welch*, $Blackman-Tukey$ και *Daniell*. Κοινός παρανομαστής σε όλες τις μεθόδους είναι η προσπάθεια να μελετηθούν και να παρουσιαστούν οι παράμετροι που χρησιμοποιήθηκαν σε κάθε μέθοδο. Αυτή η ανάγκη προέκυψε από το γεγονός πως στην υπάρχουσα βιβλιογραφία δεν αναφέρονταν οι παράμετροι που έδιναν τα αντίστοιχα αποτελέσματα σε κάθε μέθοδο. Όπως έδειξαν τα συστηματικά πειράματα της παρούσας μελέτης, η αναπαραγωγή προτεινόμενων μεθόδων με ιδιαίτερη επιμέλεια στην επιλογή των παραμέτρων οδήγησε στην επίτευξη καλύτερων αποτελεσμάτων. Αφού υπολογισθεί η πυκνότητα φάσματος ισχύος σε κάθε τεμάχιο, εξάγονται οι συχνότητες που αντιστοιχούν στο μέγιστο μέτρο της πυκνότητας φάσματος ισχύος, καθώς οι γειτονικές της. Στη συνέχεια, εφαρμόζεται τετραγωνική παρεμβολή με σκοπό την καλύτερη εκτίμηση της ΣΗΔ. Αφού γίνει η εξαγωγή της ΣΗΔ σε κάθε τεμάχιο, η προκύπτουσα

χρονοσειρά συμμετέχει στη διαδικασία ελέγχου της αυθεντικότητας του σήματος. Αυτό καθίσταται δυνατό με την αντιπαραβολή του εξαχθέντος σήματος ΣΗΔ και του σήματος ΣΗΔ αναφοράς, όπως έχει μετρηθεί από ειδικά όργανα υψίστης ακριβείας. Ο έλεγχος αυθεντικότητας πραγματοποιείται ως προς δύο κριτήρια πιστότητας. Το πρώτο κριτήριο είναι ο συντελεστής συσχέτισης ενώ το δεύτερο κριτήριο είναι η τυπική απόκλιση των σφαλμάτων μεταξύ του εξαχθέντος σήματος ΣΗΔ και του αντίστοιχου αναφοράς. Όσον αφορα το πρώτο σήμα, η αύξηση του παραθύρου απο $20\,sec$ σε $40\,sec$ δεν οδηγεί σε βελτίωση της ακρίβειας των ελέγχων. Τα καλύτερα αποτελέσματα για το συντελεστή συσχέτισης παρουσιάζονται αναλύοντας το αρχικό σήμα περί την πρώτη και τρίτη αρμονική με ποσοστό που φτάνει το 99.83% για τη μέθοδο $Welch$. Εξίσου υψηλό συντελεστή συσχέτιτσης παρουσιάζουν και οι υπόλοιπες μέθοδοι με την $Capon$ να αγγίζει το 99.77% για την τρίτη αρμονική και τις μεθόδους $ESPRIT$ και $MUSIC$ να φτάνουν το 99.79%. Αξιόλογα αποτελέσματα προκύπτουν και κατά τον έλεγχο με τη χρήση της τυπικής απόκλισης των σφαλμάτων. Η μέθοδος $Welch$ για την πρώτη και τρίτη αρμονική παρουσιάζει τη μικρότερη τυπική απόκλιση σφαλμάτων, $1.069\,mHz$. Εξίσου μικρά σφάλματα παρουσιάζουν και οι υπόλοιπες μέθοδοι, με την $ESPRIT$ και $MUSIC$ να ακολουθούν τη μέθοδο $Welch$ με τυπική απόκλιση σφαλμάτων $1.202\,mHz$ και $1.208\,mHz$, αντιστοίχως.

Όσον αφορα το δεύτερο σήμα, την ηχητική καταγραφή, η ακρίβεια για όλες τις μεθόδους ως προς το συντελεστή συσχέτισης ξεπερνά το 90%. Ο μεγαλύτερος συντελεστής συσχέτισης παρατηρείται για τις μεθόδους $Capon$, $ESPRIT$ και $MUSIC$ και ανέρχεται σε 93.18%. Η ακρίβεια ως προς το δεύτερο κριτήριο, την τυπική απόκλιση των σφαλμάτων, παρουσιάζει τις μικρότερες τιμές για τις μεθόδους $Short-Time\,Fourier\,Transform$ και $Blackman-Tukey$. Με την αύξηση του παραθύρου στα $50\,sec$ παρατηρείται αύξηση της ακρίβειας του ελέγχου αυθεντικότητας και για τα δύο κριτήρια. Όπως και στην περίπτωση των $33\,sec$, το μεγαλύτερο συντελεστή συσχέτισης, 94.44%, παρουσιάζουν οι μέθοδοι $Capon$, $ESPRIT$ και $MUSIC$. Ως προς την τυπική απόκλιση των σφαλμάτων, οι μέθοδοι $Short-Time\,Fourier\,Transform$ και $Blackman-Tukey$ δίνουν τις μικρότερες τιμές.

Το υπολογιστικό κόστος της κάθε μεθόδου αποτελεί έναν άλλο σημαντικό

παράγοντα στην επιλογή μεθόδου και παραθύρου. Οι μέθοδοι εξαγωγής και ελέγχου αυθεντικότητας της ΣΗΔ στο πρώτο σήμα με χρήση παραθύρου των $20\,sec$ απαιτούν χρόνους που ποικίλουν ανάλογα με τη μέθοδο που υιοθετείται. Οι μέθοδοι $Short-Time\,Fourier\,Transform$, $Blackman-Tukey$ και $Daniell$ απαιτούν χρόνους μικρότερους των $5\,sec$. Οι μέθοδοι $ESPRIT$ και $MUSIC$ βρίσκονται σε υψηλότερα, αλλά αποδεκτά επίπεδα, περί τα $76\,sec$ δίνοντας καλές ακρίβειες. Η μέθοδος $Capon$ απαιτεί πολύ χρόνο, $1395\,sec$, ο οποίος την καθιστά μη πρακτική σε εφαρμογές. Ίδιοι, σχεδόν, χρόνοι απαιτούνται, όταν αναλύεται το σήμα περί τη δεύτερη και τρίτη αρμονική. Οι χρόνοι υπολογισμού όταν χρησιμοποιείται παράθυρο μήκους $40\,sec$ παρουσιάζουν ανάλογη συμπεριφορά όσον αφορά τη σειρά κατάταξης των μεθόδων. Αξίζει να σημειωθεί πως σε κάθε περίπτωση παρατηρείται αύξηση του απαιτούμενου χρόνου δύο έως τρεις φορές. Οι μέθοδοι εξαγωγής ΣΗΔ και ελέγχου αυθεντικότητας στο δεύτερο σήμα απαιτούν περισσότερο χρόνο από ό,τι στο πρώτο σήμα. Και για τις δύο επιλογές παραθύρων, η μέθοδος του $Welch$ απαιτεί χρόνο που δεν ενδείκνυται για πρακτική χρήση και ανέρχεται σε $23902\,sec$. Ανάλογη χρονική απαίτηση εμφανίζει και η μέθοδος του $Capon$ που ανέρχεται σε $2248\,sec$. Οι υπόλοιπες μέθοδοι είναι γρήγορες και πρακτικώς εφαρμόσιμες.

Chapter 1

Introduction

In modern society, a huge amount of multimedia content is present in our everyday life in the form of audio and video recordings. However, the multimedia content can be edited, altered and modified for various purposes. This is the reason why a lot of frauds occur in terms of altering footage and digital content without been noticed by the victim. As Albert Einstein said, *"It has become appallingly obvious that our technology has exceeded our humanity"*. It is in this stage where forensic analysis plays a crucial role in detecting and tracking these vicious changes of digital content.

Multimedia forensic analysis is widely used due to the rapidly increased volume of shared audio and video recordings. Forensics have been studied a lot in the literature [1]. In forensic sciences, authenticating digital content and determining the time and place of recording are critical tasks. Opposed to the use of technological development for evil purposes, the advanced signal processing techniques allow us to detect if the digital material has been modified or tampered after its production. For determining the integrity and the authenticity of the digital material, few tools have been developed and used even in litigation cases, when evidence is manipulated and its originality is under question [2]. A forensic tool, which is extensively used for forgery detection in multimedia recordings and time-stamp authentication is the Electric Network Frequency (ENF) criterion, which can be used in order to determine if a recording was captured at a specific time and location [2–4].

This method is based on extracting and matching the ENF signal to a reference ENF signal from a database in order to determine whether the digital content is authentic or has been altered and estimate the time or place the recording occured.

ENF is the supply frequency in power distribution networks and its nominal value is $50\,Hz$ in Europe and $60\,Hz$ in the United States (U.S.) The major property of the ENF signal is that its value is fluctuating in a random way around its nominal one. These fluctuations vary between ± 50 to $100\,mHz$ in U.S. [5] and they are supposed to be identical through an inter-connected network.

In this thesis, we examine various spectral analysis methods for ENF estimation. All methods are applied to consecutive frames of the ENF signal recorded from the power mains as well as an audio recording. Both signals are the ones used in [6]. The fundamental ENF and its harmonics are estimated by tracking the maxima of the power spectrum after quadratic interpolation in each frame [6]. Other methods for ENF extraction, proposed in the literature, are tested using a variety of parameters during the extraction process. These tests are carried out for both the ENF signal and the speech recording, examining the peculiarities and difficulties arising from each signal and the ways to overcome them for accurate ENF extraction. Motivated by Professor Jian Li's quote "Spectral estimation is an art", here we put emphasis on the details of band-pass filtering of the raw signal prior to spectral analysis and the fine tuning of the parameters involved in spectral analysis techniques, which enable us to report more accurate results than those found in the related literature with respect to the matching of the extracted ENF and the ground truth one. The matching quality is measured with respect to correlation-coefficient and standard deviation of error. Besides the extraction of fundamental frequency, ENF extraction is carried out in the higher harmonics (e.g., the third one), which demonstrates a higher signal to noise ratio (SNR) than the fundamental one, yielding better results, as observed in [7].

This thesis tries to answer some major research questions about ENF extraction. Specifically:

⇒ Investigates the parameter tuning during the filtering process;

⇒ Conducts a systematic study of a variety of spectral estimation methods;

⇒ Provides the reader useful information and explanations about every detail and parameter of each method;

⇒ Studies the behavior of each harmonic of the ENF signal and analyzes the peculiarities of audio signal;

⇒ Measures the time requirements with respect to the method, the frame length and the filtering procedure adopted for the extraction of the ENF signal;

⇒ States every step of the procedure in order to be a useful reference for reproducing the experiments conducted.

1.1 Thesis outline

This thesis is structured in six chapters. In the first chapter a brief introduction to the thesis is presented in order to define the basic concepts.

The second chapter deals with the existing literature about the ENF extraction methods and algorithms. Also, a basic theory of the ENF is presented in order to make clear the way this signal works. Moreover, the ENF criterion is presented as basis of all methods.

A detailed presentation of the algorithms developed in the thesis and their mathematical backround is extensively discussed in the third chapter.

The fourth chapter analyses the data used in the thesis and the procedure followed for ENF extraction. Techniques are also presented for matching the extracted ENF with the reference data in order to indicate whether the recordings have been altered or not. The setup and the parameters used in all methods are also discussed.

The fifth chapter presents the evaluation of all experiments and discusses the experimental results in detail.

The last chapter concludes the thesis and indicates topics of future research.

Chapter 2

Electric Network Frequency (ENF)

2.1 Literature Overview

An introduction to ENF analysis, principles, procedures, and real world examples can be found in the seminal work of Catalin Grigoras [2–4]. The unique property of ENF signal, acting as a fingerprint, is extensively analysed, methods for obtaining the ENF, matching the ENF with a ground truth signal, and in-depth understanding of its behaviour are presented.

A procedure for ENF extraction, using Short-Time Fourier Transform (STFT) followed by a quadratic interpolation, and a mean square error metric for ENF matching is proposed in [8, 9]. Adaptive techniques for ENF extraction based on dynamic programming are presented in [6], where a detailed comparison between various techniques is made for digital audio recordings. In [5], different parametric and non-parametric methods of ENF estimation are elaborated, addressing the problem of extracting location information from the ENF signal. A more precise and detailed study, focusing on determining the intra-grid location of recordings, is discussed in [10].

Many studies focus on time verification of recordings using different matching procedures, such as the correlation coefficient approach proposed in [11]. In [12], the ENF signal is modeled as an autoregressive process and a decor-

relation based approach is adapted to ENF matching. A computationally efficient form of maximum-likelihood estimation is presented in [13] via a multitone harmonic model, where signal-to-noise ratio (SNR) is also considered in the process. Cramer-Rao bound is, also, used for error estimation. The maximum-likelihood frequency estimation is given by $\hat{\omega}_0^{ML} = \text{argmax}_{\omega_0} \|A(\omega_0)\mathbf{y}\|^2$, where A is a projection matrix. In [14], a novel approach for authenticating audio signals is proposed. Max offset for cross-correlation between the extracted signal and the reference ENF is adopted. Besides answering whether the digital content was edited, this method, also, determines the location the alterations occur. In [15], ENF signal is estimated by making use of multiple harmonics with a weighted summation of spectral bands according to local SNR. The estimated spectrum can be formulated as: $S(f) = \sum_{k=1}^{L} w_k P_{B,k}(kf)$, where $P_{B,k}(kf)$ is the power spectrum for a given time-frame and w_k is the weight for the harmonic bands. The calculation of w_k is based on the SNR.

Apart from digital audio recordings, the ENF signal can be extracted from digital video content recorded in indoor environments with the presence of fluorescent lighting [16,17] in order to estimate the time of recording and verify its authenticity. In [18], two ENF extraction methods for video recordings are developed, exploiting the rolling shutter mechanism of an imaging sensor. Different pixel positions in consecutive frames and static regions are detected for each method, respectively in order to derive the ENF flactuations.

2.2 ENF properties

Electric networks operate at a specific frequency, which is $50\,Hz$ in Europe and $60\,Hz$ in the US. This frequency is called Electric Network Frequency (ENF). ENF is not stable and it flactuates around its nominal value. These flactuations occur instantly due to the differences between the produced and consumed electrical power [2]. Every moment the need for electrical power differs and thus the rotational speed of energy generators changes. These unbalances cause the ENF to vary around $50\,Hz$ and $60\,Hz$.

ENF flactuations are non-periodic and they can not be predicted even if

we have recorded for a long period of the signal. Although the ENF pattern is not predictable through time, it is commonly accepted that its behaviour is stable in the whole network, so the pattern does not change [3]. This makes the ENF signal identical for authenticating digital content in case a reference database of ENF signals is captured. The ENF can be formulated at any given moment as [2]:

$$f = [50 \pm \Delta f]\, Hz \tag{2.1}$$

where Δf is the deviation between the instantaneous frequency and its nominal value. In Figure 2.1, a $40\, min$ recording of the ENF pattern is shown. This ENF signal was captured by a Frequency Disturbance Recorder (FDR) in the US. ENF flactuates around its fundamental frequency of $60\, Hz$.

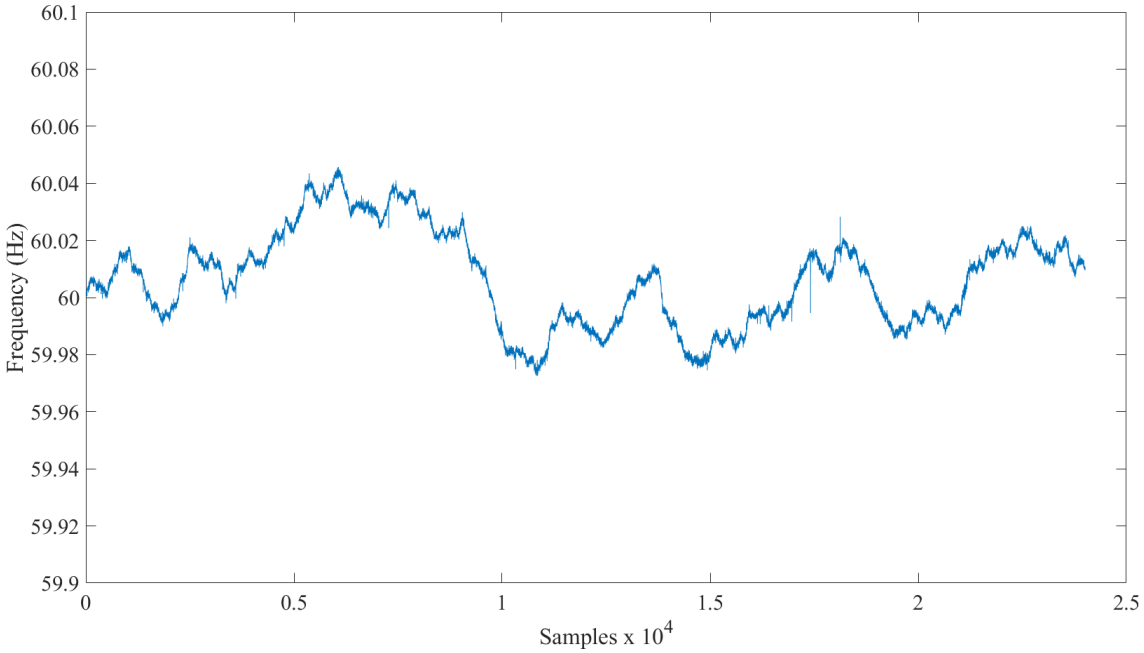

Figure 2.1: Reference ENF pattern

According to [3], operation conditions of ENF vary with respect to Δf value:

$\Rightarrow$ If $\Delta f \leq 50\, mHz$, the flactuations are considered to be normal.

$\Rightarrow$ If $50\,mHz < \Delta f \leq 150\,mHz$, the flactuations exceed the normal values, but no danger of corrupting the ENF exists.

$\Rightarrow$ If $\Delta f > 150\,mHz$, the flactuations yield major disturbances and the network proper functioning is at risk.

ENF unique properties make it a powerful tool in forensic analysis. Recent advances in signal processing will enable scientists to further investigate new ways and methods for accurately determining ENF signal. The basic properties of the ENF signal are listed below [3]:

$\Rightarrow$ ENF signal occurs in random way. No patterns exist.

$\Rightarrow$ Its pattern is identical in the same network.

$\Rightarrow$ Apart from the fundamental frequency, ENF also exists in its higher harmonics [7].

2.3 The ENF criterion

The ENF criterion is a powerful tool for authenticating digital media content. It is introduced by Catalin Grigoras [2, 3, 9] for determining the time and location of digital recordings and it is based on the unique properties of the ENF signal.

ENF flactuations leave a unique timestamp on the digital recordings. Given the fact that these recordings are exactly the same at any point on the electrical grid at the same time, it is clearly stated that their analysis could provide useful information about the time of the recording and their integrity. Traces of these deviations exist in the digital recordings, when the recordings are captured by devices connected to the grid. Such disturbances have to be isolated in order to be extracted in the most accurate way. After the extraction, the ENF should be matched against a reference grid ENF signal at the corresponding time.

Both the extraction methods and the matching algorithms are investigated in the most challenging environments, where noise is present in the

recordings. The subsequent chapters deal with different extraction methods, trying to provide the reader an in-depth insight of the different extraction approaches, using state-of-the-art methods.

Chapter 3

ENF Procedures

ENF extraction can be achieved through a variety of different paths according to the needs and the available algorithms. Based on the properties of ENF signal and electrical networks, three extraction methods exist, as presented by Grigoras in [2] and [3]. These methods can be categorized as follows:

⇒ Time/frequency domain analysis: It is based on spectrogram computation. The matching procedure is visually accomplished against the reference signal. It is suitable for short-time recordings (usually no longer than $10 - 15$ min).

⇒ Frequency-domain analysis: In this case, the periodogram is applied to short-time segments of data and the local maxima of the spectrum magnitude are located. The corresponding frequency of the local maximum is the estimate of ENF. The estimated ENF is compared to the reference ENF entry of a ground truth database. This method is used after a band-pass filter is applied to the raw data. Other spectral methods can be applied as well.

⇒ Time-domain analysis: It uses the zero-crossing measurements after applying a band-pass filter around the nominal frequency.

An automated ENF extraction procedure was proposed by Cooper in [8, 9]. It is presented and discussed in detail next. The matching part of

the extracted ENF signal against the reference one is conducted with respect to the correlation-coefficient approach [11] or a slight modification of mean square error, which uses the standard deviation of errors, as done in [6].

3.1 Dataset description and experimental setup

The two datasets used in [6] and the ENF ground truth associated to them are discussed here. In particular, the first dataset was recorded by connecting an electric outlet directly to the internal sound card of a desktop computer with a voltage divider and the second one was a speech recording captured by the internal microphone of a laptop computer. Both recordings were sampled at 44.1 kHz using 16 bits per sample.

Two sets of parameters are used in this thesis, as shown in Table 3.1. The first set of frame length L_1 is the same as in [6]. Afterwards, the frame length L_2 is increased in order to demonstrate its impact in accuracy and computational complexity. A time shift of 1 *sec* equals to 441 *samples* of the raw data as long as the sampling frequency is equal to 441 Hz (after downsampling).

Table 3.1: Frame parameters (in sec)

Parameters	*Data* 1 (recorded from power mains)	*Data* 2 (speech recording)
Time shift, T	1	1
Frame length, L_1 [6]	20	33
Frame length, L_2	40	50

3.2 ENF extraction procedure

The procedure followed for ENF extraction is proposed by Cooper in [8, 9] and is slightly modified according to the needs of this work. The basic steps of the procedure entail: a) downsampling the raw signal; b) band-pass filtering,

which depends on the nominal frequency of the ENF signal; c) extraction procedure and corresponding algorithms; d) quadratic interpolation calculations for accurate frequency estimation and e) matching procedure against the reference database, which contains the ground-truth ENF measurements. The whole process is detailed not only to guarantee a self-contained treatement of the topic investigated, but also to enable reproducibility of the experiments conducted by the interested reader.

Firstly, the original recordings are down-sampled to a frequency that contains the fundamental frequency and some of its higher harmonics, i.e., $F_s = 441\,Hz$. By doing so, the ENF frequency is preserved into the down-sampled signal and the computational cost is significantly reduced. The initial sample decimation depends on the frequency of interest, i.e., $50\,or\,60\,Hz$ and the number of harmonics to be studied. In this study, the first three harmonics are to be studied and the nominal frequency is $60\,Hz$.

The second step is a band-pass filtering of the signal around the nominal ENF or its harmonics. Signal filtering is of high importance and filter parameter selection should be done very carefully. In contrast to Data 1, speech recording (Data 2) contains many interferences and its SNR is low. The first and the third harmonics are too weak to be estimated [6]. When the second harmonic of Data 2 is computed, some interesting properties are arised. They are discussed in detail. The procedure of signal filtering requires two basic specifications. a) The band-pass edges and b) the filter order. The filter order should be an odd number. In our work, as concerns Data 1, the band-pass edges of the filter are set at $59.9\,Hz$ and $60.1\,Hz$ for the fundamental frequency and the filter order is 1501. The second harmonic band-pass edges are set at $119.9\,Hz$ and $120.1\,Hz$, respectively. The filter order remains the same, i.e., 1501. The third harmonic band-pass edges are set at $179.9\,Hz$ and $180.1\,Hz$ and the filter order is chosen equal to 1001. Any changes in band-pass edges or filter order did not have any impact on the results of matching procedure.

As mentioned before, the ENF signal apart from its nominal frequency, can also be found in the second, third, etc. harmonics. In Figures 3.1, 3.2, and 3.3, the Power Spectral Density (PSD) of the first three harmoncis is

depicted, after band-pass filtering applied to Data 1. In each case, we can attest that the filtering was correctly applied by observing the magnitude of the PSD of the band-passed ENF signal and the PSD of the original (raw) signal.

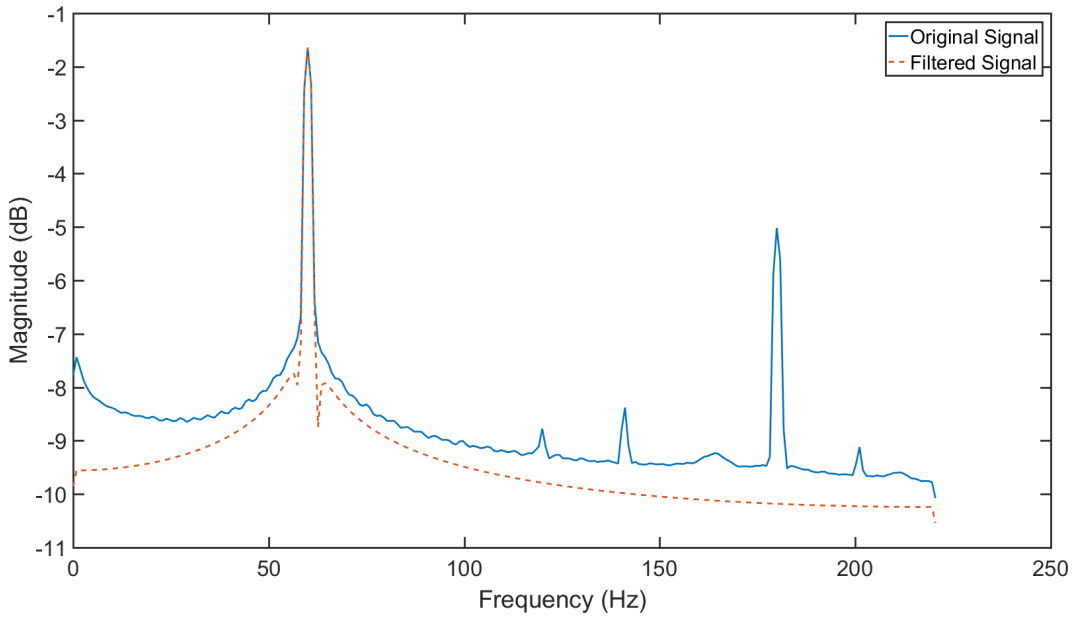

Figure 3.1: PSD of the signal before and after band-pass filtering around the first harmonic (i.e., $60\,Hz$).

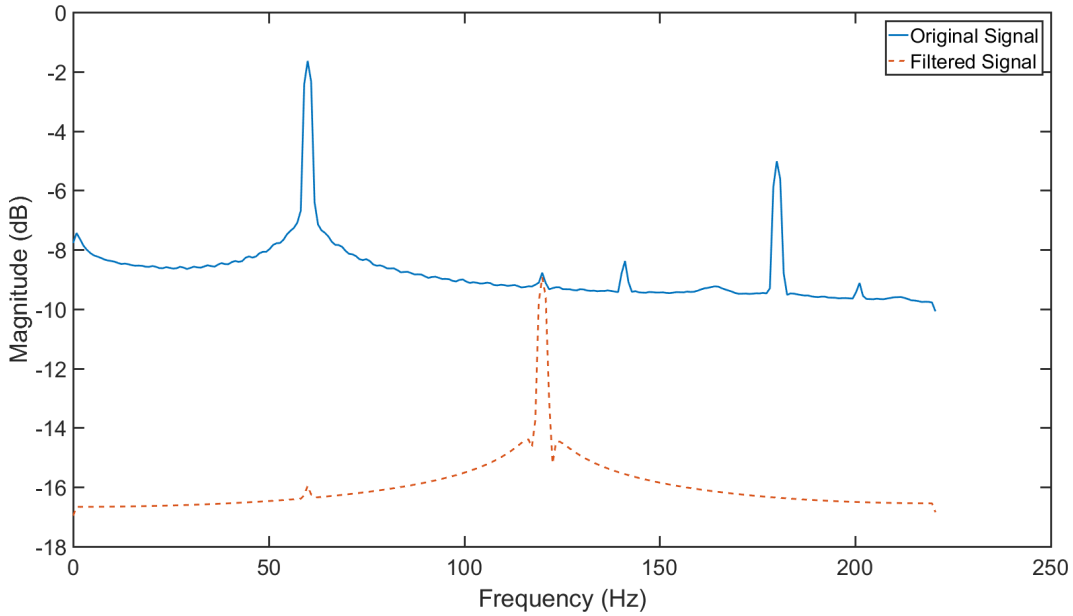

Figure 3.2: PSD of the signal before and after band-pass filtering around the second harmonic (i.e., $120\,Hz$).

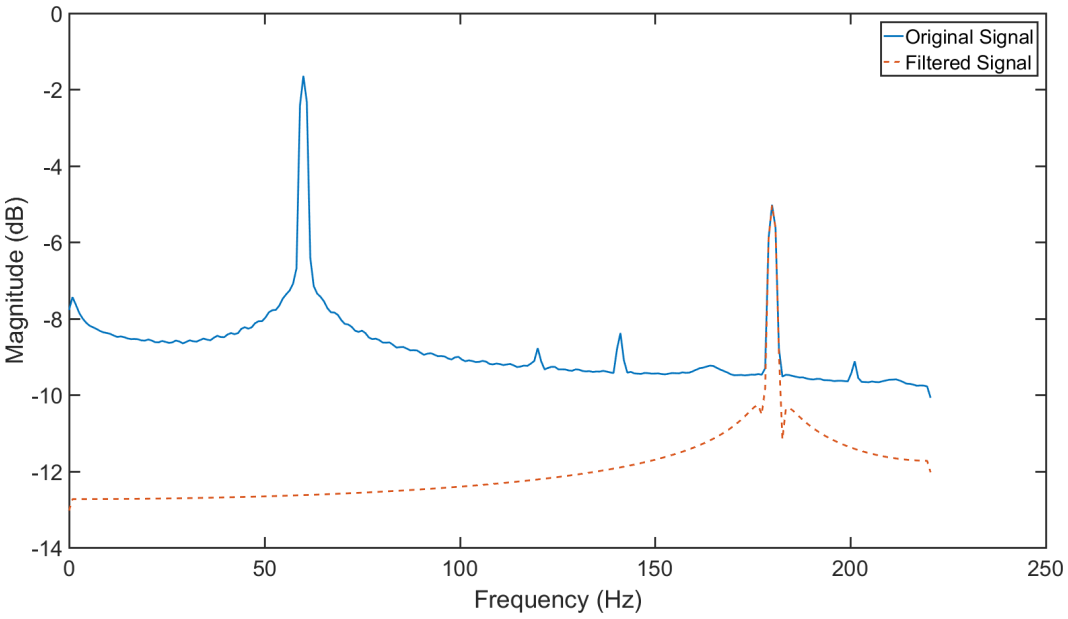

Figure 3.3: PSD of the signal before and after band-pass filtering around the third harmonic (i.e., $180\,Hz$).

In each case, a Hamming window with length equal to filter order is used in band-pass filter design. The aforementioned band-pass filter specifications are adopted after a wide range of experiments. Fine tuning of the filter parameters is found to be of crucial importance. Proper parameter selection yields accurate ENF extraction, while a slight parameter modification yields less accurate results.

During the next step, the filtered signal is split into K overlapping frames as shown in Figure 3.4. Each frame is obtained by applying a rectangular window of length L seconds to the filtered signal and is shifted by $T = 1$ sec from its immediate predecessor frame. Two choices for L, denoted as L_1 and L_2, are indicated in Table 3.1 along with T. So, for the aforementioned frame parameters, the frame size would be $N = frame\,length \times sampling\,frequency$, i.e. $[20, 40, 33, 50] \times 441\,samples$. Frames are common in short-term signal processing due to the fact that they increase frequency resolution and improve ENF estimation.

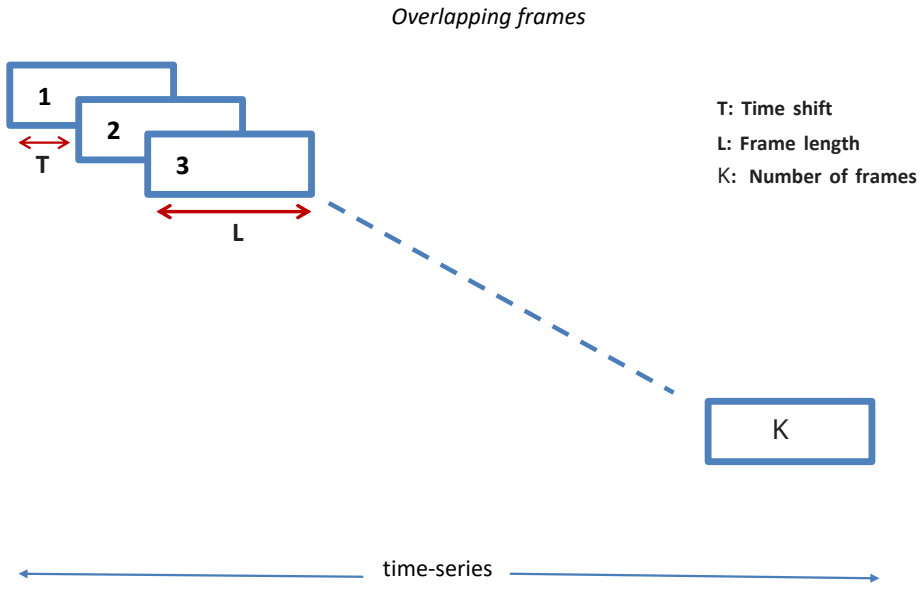

Figure 3.4: Overplapping windows for spectral estimation.

The next step involves spectral estimation. For each consecutive frame, the power spectrum is estimated by various spectral analysis techniques. This step is of high importance. Each method applied has its own advantages and disadvantages. So, a careful choice should be done. In this thesis, a wide range of different methods are tested in order to demonstrate the benefits of each approach. A systematic study is conducted using a refined periodogram, a refined filter-bank, and high-resolution spectral estimation methods. These approaches are discussed in detail in Chapter 4.

Regarding Data 1, the band-pass filtering was crucial for proper and accurate ENF extraction. In Figure 3.5, the extracted ENF signal by the Capon method for the first harmonic is shown. For illustration purposes the estimated ENF signal is shifted by $0.05\,Hz$.

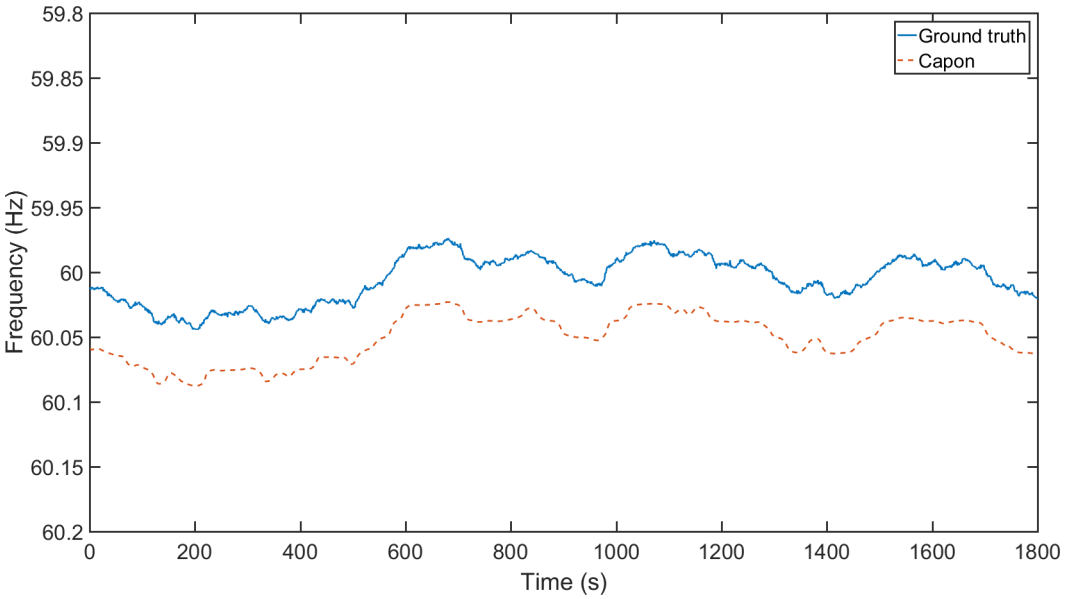

Figure 3.5: ENF signal of the band-pass filtered Data 1 signal around the first harmonic. (filter order 1501, band-pass edges 59.95 Hz and 60.05 Hz). Capon method was used for the extraction of the ENF.

The speech recording (Data 2) presented a completely different behavior. This is because of the noise existing in this kind of recordings due to the equipment used to capture them. Using the same filter order as previously, the results of ENF using the Short-Time Fourier Transform (STFT) approach, showed a completely false diagram. For given band-pass edges at 119.95 Hz and 120.05 Hz and filter orders 1501 and 4801 respectively, it is clear that a false choice of filter order is going to lead to false results, as shown in Figures 3.6 and 3.7.

In Figure 3.6, the extracted ENF for audio signal (Data 2) is totally false due to interferences that were not cut during the filtering step despite the strict cut-off frequencies. These results arose due to the low filter order. Second harmonic of Data 2, after band-pass filtering with filter order 1501 and band-pass edges 119.95 Hz and 120.05 Hz. The STFT was used for the extraction of the second harmonic. The second harmonic was scaled by a factor 2 to fall in the range of ENF.

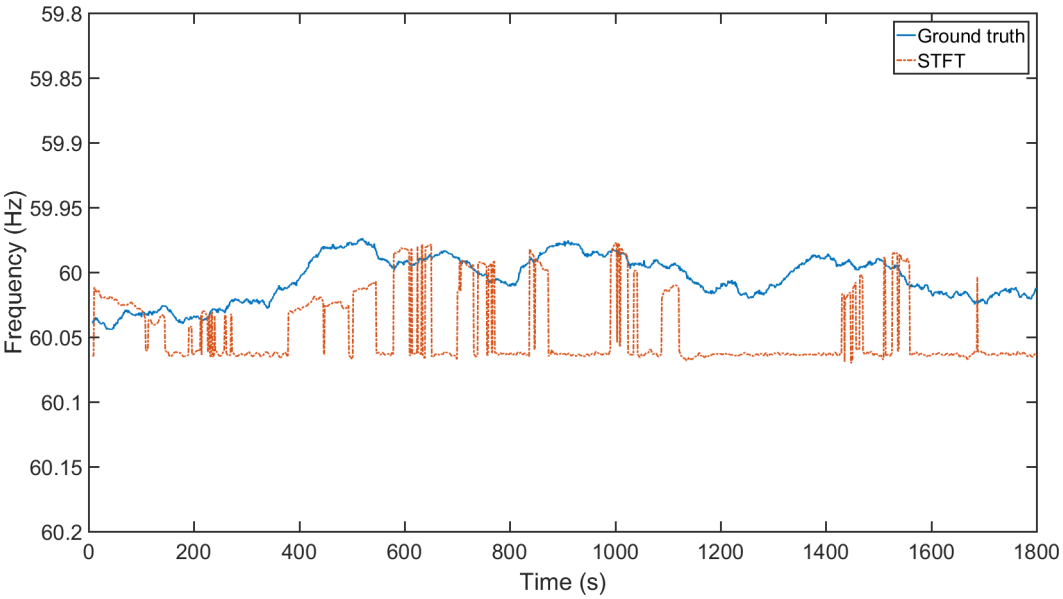

Figure 3.6: ENF signal of Data 2 with filter order 1501 for the second harmonic with band-pass edges 119.95 Hz and 120.05 Hz. The STFT method was used for the ENF extraction.

In Figure 3.7, the extracted ENF for audio signal (Data 2) is clearly improved and provided accurate ENF extraction results. Second harmonic of Data 2, after band-pass filtering with filter order 4801 and band-pass edges 119.95 Hz and 120.05 Hz. The STFT was used for the extraction of the second harmonic. The second harmonic was scaled by a factor 2 to fall in the range of ENF.

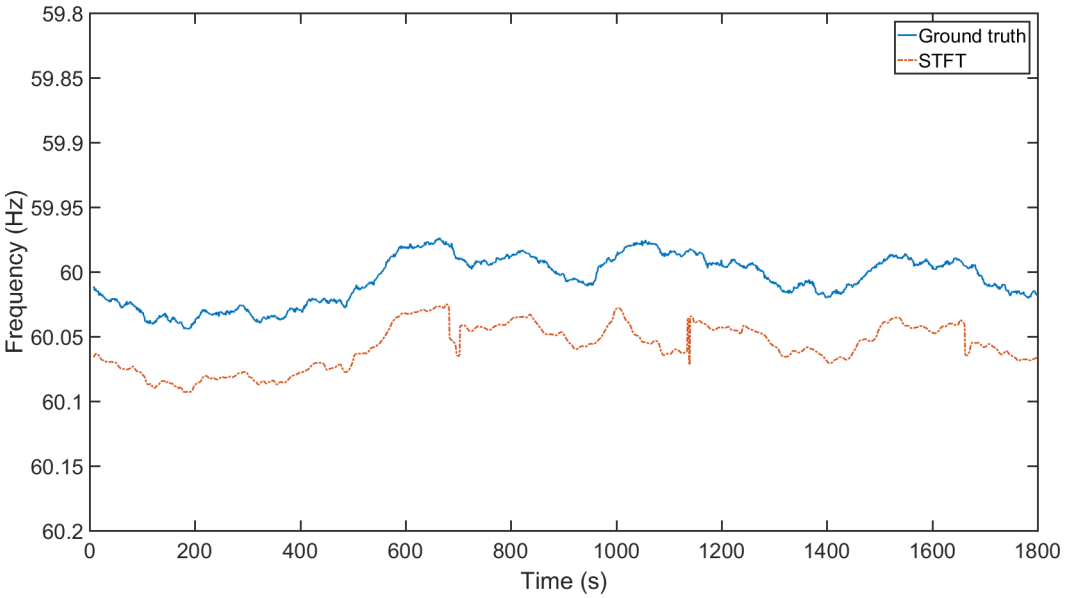

Figure 3.7: ENF signal of Data 2 with filter order 4801 for the second harmonic with band-pass edges 119.95 Hz and 120.05 Hz. The STFT method was used for the ENF extraction. ENF signal is shifted by 0.05 Hz for illustration puproses.

In Figure 3.8, the ESPRIT method was used for ENF extraction for the second harmonic of Data 2 using the proper filter order (i.e., 4801). The ESPRIT method provides an ENF signal visually closer to the reference one than the STFT in Figure 3.7. Second harmonic of Data 2, after band-pass filtering with filter order 4801 and band-pass edges 119.95 Hz and 120.05 Hz. The ESPRIT was used for the extraction of the second harmonic. The second harmonic was scaled by a factor 2 to fall in the range of ENF.

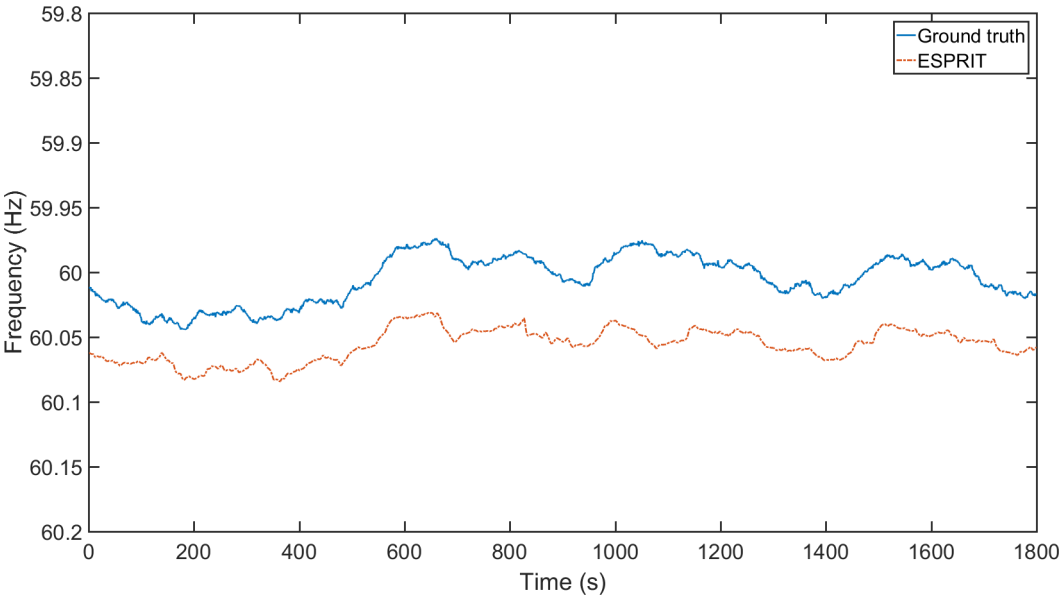

Figure 3.8: ENF signal of Data 2 with filter order 4801 for the second harmonic with band-pass edges 119.95 Hz and 120.05 Hz. ESPRIT method was used for the extraction. The ENF extracted by ESPRIT is shifted by 0.05 Hz for illustration purposes.

The fine tuning of filtering parameters is of vital significance for the accurate ENF extraction. The first example (Figure 3.6) leads to a matching accuracy of 2%, while the second one (Figure 3.7) leads to an accuracy of 92%. These results would be discussed in detail in Chapter 5.

A comparison was made, also, using the thrid harmonic of the speech recording, which is subsequently scaled by a factor 3 to fall in the range of the ENF. As mentioned in [6], the third harmonic is too weak to provide an efficient estimation of the ENF. Even if the results are rather rough, we can also point out the differences between the ENF signal extracted using a low filter order and a high one. The order of 4801 provides a smoother curve, whose values are closer to the real frequency values. Figures 3.9 and 3.10 demonstrate these differences. It is worth mentioning that the band-pass edges and the extraction method are those used for the second harmonic, previously.

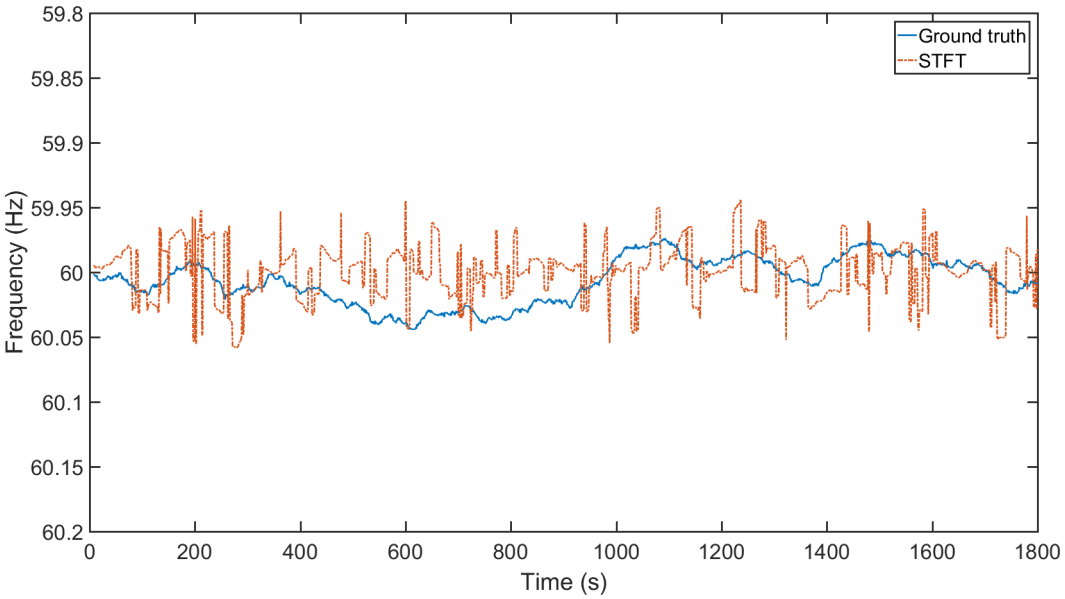

Figure 3.9: ENF signal extracted from Data 2 filtered with filter order 1501 and band-pass edges 179.95 Hz and 180.05 Hz. The third harmonic was scaled by a factor 3 to fall within the ENF range.

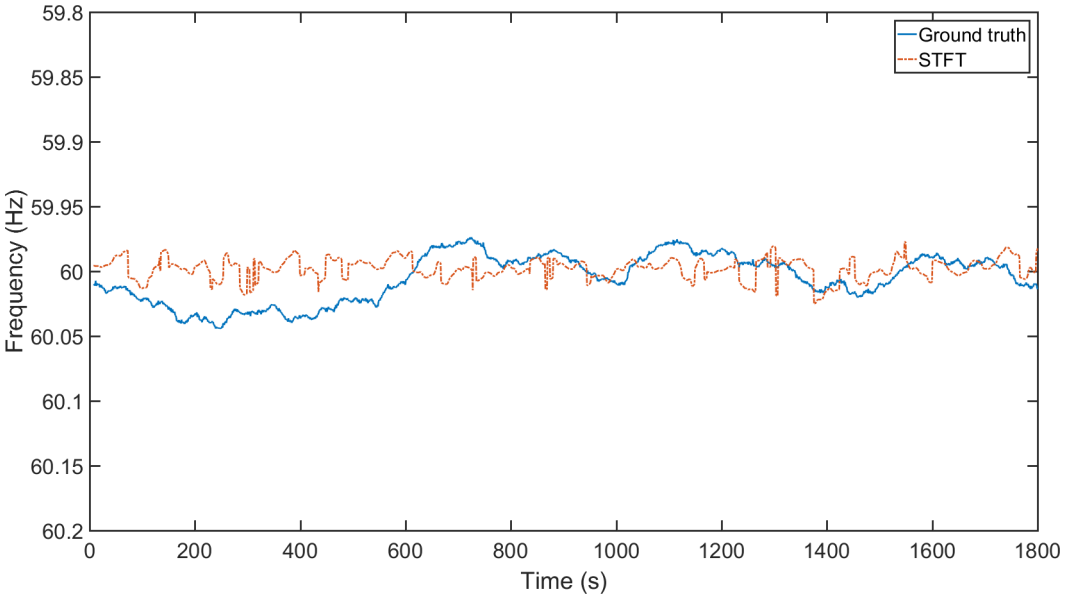

Figure 3.10: ENF signal extracted from Data 2 filtered with filter order 4801 and band-pass edges 179.95 Hz and 180.05 Hz. The third harmonic is scaled by factor 3 to fall within the ENF range.

3.3 Quadratic interpolation

Let $\hat{\phi}_r(\omega_\ell)$ be the periodogram of $N = L\,F_s$ samples-long rth frame, where $\omega_\ell = \frac{2\pi}{N}\ell$, $\ell = 0, 1, \ldots, N-1$, are the frequency samples and F_s is the sampling frequency of each signal. The frequency sample $\omega_{\ell_{\max}}$, which corresponds to the maximum periodogram value is extracted as a first ENF estimate.

In order to obtain a more accurate estimation, a quadratic interpolation (QI) is employed, which fits a quadratic model to the logarithm of the estimated power spectrum about $\omega_{\ell_{\max}}$ [6, 8, 19]. To estimate $\omega_{\ell_{\max}}$, one estimates the power spectral density of each frame and searches for its maximum magnitude. Power spectrum is calculated for each frame. Power spectrum is equal to the normalized squared magnitude of the discrete-time Fourier transform of each frame. The frame length is used for normalization. QI offers a low computational cost and enables the extraction of the ENF signal with high resolution. The steps of QI are briefly described next as in [9]. Let $\beta_q = \log \hat{\phi}_r(\omega_{\ell_{\max}+q})$, $q = -1, 0, 1$:

$\Rightarrow$ Select bin $\beta_0 = \log \hat{\phi}_r(\omega_{\ell_{\max}})$

$\Rightarrow$ Select the adjacent bins on either side of β_0, i.e., $\beta_{-1} = \log \hat{\phi}_r(\omega_{\ell_{\max}-1})$ and $\beta_1 = \log \hat{\phi}_r(\omega_{\ell_{\max}+1})$;

$\Rightarrow$ Fit a quadratic model to these three points;

$\Rightarrow$ Find the interpolated value of the quadratic model, which corresponds to quadratic peak δ;

By using the two adjacent frequencies around $\omega_{\ell\max}$, a more accurate ENF estimate is obtained as $\omega = \omega_{\ell_{\max}} + \delta$, where

$$\delta = \frac{1}{2}\frac{\beta_{-1} - \beta_1}{\beta_{-1} - 2\beta_0 + \beta_1}(\omega_{\ell_{\max}+1} - \omega_{\ell_{\max}}) \tag{3.1}$$

The frequency estimated by the QI is stored as the extracted ENF value. Hereafter, the aforementioned spectral estimation method is replaced by other non-parametric and parametric spectral analysis methods.

3.4 Matching procedure

Having extracted the ENF from either the recorded signal from the power mains (Data 1) or the audio recording (Data 2), a matching procedure has to be performed against the ground truth information in order to identify the time the recording has been captured. Using the notation introduced in [6], let $\mathbf{f} = [f_1, f_2, \ldots, f_K]^T$ be the extracted ENF signal, which comprises the ENF estimated at each second. Let also $\mathbf{g} = [g_1, g_2, \ldots, g_{\tilde{K}}]^T$ for $\tilde{K} > K$ be the reference ground truth ENF, which comprises the actual ENF values monitored and kept in the records of a power corporation at various time instants. In [8], the association is being done by minimizing the squared error between $\mathbf{f}$ and $\tilde{\mathbf{g}}(l) = [g_l, g_{l+1}, \ldots, g_{l+K-1}]^T$, i.e.,

$$l_{opt} = \underset{l=1}{\overset{\tilde{K}-K}{\operatorname{argmin}}} \|\mathbf{f} - \tilde{\mathbf{g}}(l)\|_2^2 \tag{3.2}$$

An alternative matching criterion, proposed in [11], is the correlation matching, i.e.,

$$l_{opt} = \underset{l=1}{\overset{\tilde{K}-K}{\operatorname{argmax}}}\, c(l) \tag{3.3}$$

where $c(l)$ is the sample correlation coefficient between $\mathbf{f}$ and $\tilde{\mathbf{g}}(l)$ defined as

$$c(l) = \frac{\mathbf{f}^T \tilde{\mathbf{g}}(l)}{\|\mathbf{f}\|_2 \, \|\tilde{\mathbf{g}}(l)\|_2} \tag{3.4}$$

To measure the accuracy of ENF extraction by various algorithms, one may employ the maximum correlation coefficient $c(l_{opt})$ measured on either Data 1 or Data 2, respectively. Alternatively, one may employ the standard deviation of the error between the true ENF and the estimated one. The former figure of merit was found to be more accurate than the latter one [11]. A flowchart depicting the process implemented is presented in Figure 3.11.

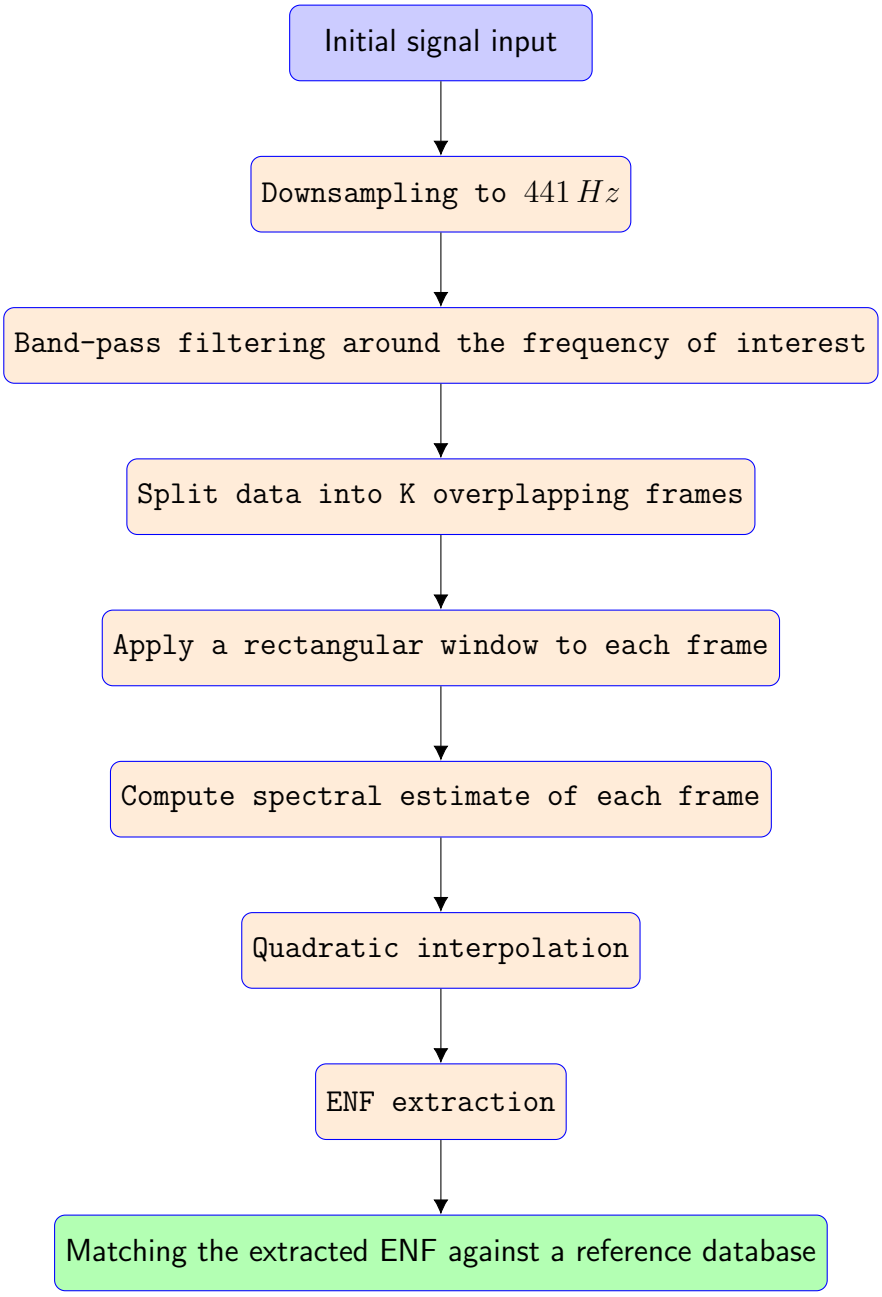

Figure 3.11: Flowchart of ENF extraction procedure.

Chapter 4

Algorithms for Spectral Analysis

4.1 Short-time Fourier transform

The Short-Time Fourier Transform (STFT) is one of the most common algorithms used in ENF extraction. STFT is based on dicrete-time Fourier transform and is computed by means of the Fast-Fourier Transform (FFT). It constitutes a basic and powerful tool in audio signal processing and also, in analyzing quasi-stationary signals [20]. The initial recorded signal (time-domain) is divided into overlapping frames multiplied by a slidding window $w()$. STFT is defined [20]:

$$X_n(\omega_k, m) = \sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-j\omega_k m} \tag{4.1}$$

The periodogram is calculated by squaring the magnitude of STFT. It is a basic tool in forensics analysis. Spectrogram visually respresents the frequency spectrum of signal as it varies within time.

4.2 Welch method

A refined periodogram method is the Welch method [21]. In this method, each frame is divided into overlapping segments and each segment is multiplied by a temporal window. Let $u(t)$ and $y_j(t)$ denote the temporal window and jth segment, respectively. Here, adjacent segments overlap by 1000 samples and each segment has length of $M = \frac{N}{4} = \frac{LF_s}{4}$ samples. The Welch estimate of power spectral density (PSD) is given by

$$\hat{\phi}_w(\omega) = \frac{1}{S} \sum_{j=1}^{S} \hat{\phi}_j(\omega) \tag{4.2}$$

where $S = 7$ and $\hat{\phi}_j(\omega)$ is the windowed periodogram corresponding to $y_j(t)$, i.e.,

$$\hat{\phi}_j(\omega) = \frac{1}{MP} \left| \sum_{t=1}^{M} u(t)\, y_j(t)\, e^{-i\omega t} \right|^2 \tag{4.3}$$

with P denoting the power of the temporal window $u(t)$. A rectangular window has been employed. The Welch method yields accurate ENF estimation without being affected by interferences, especially in the second harmonic of the ENF in both datasets. In speech dataset, the segment length equals to $M = \frac{N}{2} = \frac{LF_s}{2}$ samples.

4.3 Blackman-Tukey spectral estimator

The Welch estimator (4.2) can be related to the Blackman-Tukey (BT) spectral estimator for suitable choices of the lag window and the autocovariance estimate [21]. The tradeoff between spectral resolution and statistical variance should be considered in order to choose the window's length. Accordingly, a natural choice for a refined periodogram is the Blackman-Tukey estimate given by

$$\hat{\phi}_{BT}(\omega) = \sum_{k=-(M-1)}^{M-1} w(k)\, \hat{r}(k)\, e^{-i\omega\, k} \tag{4.4}$$

where $M = \frac{N}{2} = L\,\frac{F_s}{2}$ for the first and third harmonic and $M = N = L\,F_s$ for the second harmonic in both datasets.

4.4 Daniell method

Another non-parametric method is the Daniell method [21], which yields the refined spectral estimate

$$\hat{\phi}_D(\omega_\ell) = \frac{1}{2J+1} \sum_{j=k-J}^{k+J} \hat{\phi}_p(\omega_j) \tag{4.5}$$

for dense frequency samples $\omega_\ell = \frac{2\pi}{\tilde{N}}\,\ell$, $\ell = 0, 1, \ldots, \tilde{N}$. Here, the values $J = 2$ and $\tilde{N} = 4\,N = 4\,L\,F_s$ have been used. PSD estimation using Daniell method yields a non-negative function. It is worth noting that Blackman-Tukey spectral estimates are not necessarily non-negative.

4.5 Capon method

The periodogram can be interpreted as a filter bank approach, which uses a band-pass filter whose impulse response vector is given by the standard Fourier transform vector $\left[1, e^{-i\omega}, \ldots, e^{-i(N-1)\omega}\right]^T$. The Capon method, is another filter bank approach based on a data-dependent filter [21]:

$$\mathbf{h} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{a}(\omega)}{\mathbf{a}^*(\omega)\,\hat{\mathbf{R}}^{-1}\,\mathbf{a}(\omega)} \tag{4.6}$$

where $\mathbf{a}(\omega) = [1, e^{-i\omega}, \ldots, e^{-i\,m\,\omega}]^*$ and $[\cdot]^*$ denotes conjugate transposition. In (4.6), $\hat{\mathbf{R}}$ is an estimate of the auto-covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N-m} \sum_{t=m+1}^{N} \begin{bmatrix} y(t) \\ \vdots \\ y(t-m) \end{bmatrix} \Big[y^*(t), \ldots, y^*(t-m) \Big]\,. \tag{4.7}$$

Here, $m = 2$ and $N = L\,F_s$. The Capon spectral estimate is given by:

$$\hat{\phi}(\omega) = \frac{m+1}{\mathbf{a}^*(\omega)\,\hat{\mathbf{R}}^{-1}\,\mathbf{a}(\omega)}. \tag{4.8}$$

In practice, the Capon method has been found to be able to resolve fine details of a PSD, such as closely spaced peaks [21], making it a superior alternative of periodogram-based methods. Accordingly, it is a suitable method for ENF estimation.

4.6 ESPRIT

ENF estimation can be cast as an estimation of a line spectrum. Therefore, one may choose a suitable parametric method for solving the just described problem. Estimation of Signal Parameters by Rotational Invariant Techniques (ESPRIT) is a straightforward choice, as was done in [6]. In particular, one has to choose the size m of the estimated auto-covariance matrix $\hat{\mathbf{R}}$ and the dimension n of the diagonal matrix $\mathbf{D} = \mathrm{diag}(e^{-i\,\omega_1}, \ldots, e^{-i\,\omega_n})$. Let $\mathbf{I}_{m-1}$ denote the identity matrix of size $(m-1) \times (m-1)$. The frequencies $\{\omega_k\}_{k=1}^{n}$ are estimated as $-\arg(\hat{v}_k)$, where $\{\hat{v}_k\}_{k=1}^{n}$ are the eigenvalues of the estimated matrix $\hat{\boldsymbol{\phi}}$ [21]:

$$\hat{\boldsymbol{\phi}} = (\hat{\mathbf{S}}_1^*\hat{\mathbf{S}}_1)^{-1}\,\hat{\mathbf{S}}_1^*\hat{\mathbf{S}}_2. \tag{4.9}$$

In (4.9), $\hat{\mathbf{S}}_1 = [\mathbf{I}_{m-1}|\mathbf{0}]\,\hat{\mathbf{S}}$, $\hat{\mathbf{S}}_2 = [\mathbf{0}|\mathbf{I}_{m-1}]\,\hat{\mathbf{S}}$, and $\hat{\mathbf{S}}$ is the matrix having as columns the n principal eigenvectors of $\hat{\mathbf{R}}$. Here, $m = 4$ and $n = 2$. ESPRIT provides more accurate frequency estimates than other parametric methods for line spectra [21].

4.7 Multiple signal classification

The Multiple Signal Classification (MUSIC) algorithm is frequently used for frequency estimation. Thus, it is a suitable method for ENF extraction [21]. The MUSIC algorithm is used for time-delay estimation and is considered to

be a high-resolution method [22]. Its limitations exist due to the fact that model order needs to be specified before the analysis and parameter searching requires increased computing power. The first step for frequency computation involves the estimation of signal's covariance matrix $\hat{\mathbf{R}}$. Afterwards, its eigen-analysis resulting to the computation of its eigenvectors:

$$\hat{R} = \frac{1}{N} \sum_{t=m}^{N} \tilde{y}(t)\tilde{y}^*(t) \tag{4.10}$$

The frequency estimate is given by the following equation [21]:

$$P_{MU} = \frac{1}{\mathbf{a}^*(\omega)\hat{\mathbf{G}}\hat{\mathbf{G}}^*\mathbf{a}(\omega)} \tag{4.11}$$

where P_{MU} is called "pseudospectrum", because it does not represent the true PSD, but it reveals that sinusoidal components are present in the signal. $\hat{\mathbf{G}}$ denotes the matrix made from the eigenvectors $\{\hat{g}_1, \cdots, \hat{g}_{m-n}\}$ of $\hat{\mathbf{R}}$ and spans the subspace of noise. The results of the MUSIC method are similar to the ESPRIT one and led to an accurate ENF estimation.

4.8 Maximum-Likelihood estimation

The Maximum-Likelihood estimation (MLE) approach for ENF extraction is presented in [13]. The log-likelihood function's maximization of the vector $\mathbf{y}$ with respect to ω_0 results to the MLE of ω_0:

$$\hat{\omega}_0^{ML} = \underset{\omega_0}{\operatorname{argmax}} \left\| \mathbf{P}_A(\omega_0)\, \mathbf{y} \right\|^2 \tag{4.12}$$

where vector $\mathbf{y}$ contains the initial signal, and

$$P_A(\omega_0) = \mathbf{A}(\omega_0)\left[\mathbf{A}^T(\omega_0)\mathbf{A}(\omega_0)\right]^{-1}\mathbf{A}^T(\omega_0) \tag{4.13}$$

is the projection matrix onto the subspace spanned by the column of $\mathbf{A}(\omega_0)$ with dimensions $N \times (2M+1)$. The 1st row of $\mathbf{A}(\omega_0)$ is

$\left[1, cos(\omega_0 t_1), \ldots, cos(\omega_0 M t_1), sin(\omega_0 t_1), \ldots, sin(\omega_0 M t_1)\right]$.

The equation 4.12 of the MLE of ω_0 can be rewritten as

$$\hat{\omega}_0^{ML} = \underset{\omega_0}{\operatorname{argmax}}\, \mathbf{y}^T \mathbf{P}_{\mathbf{A}}(\omega_0)\mathbf{y} \tag{4.14}$$

Also, the expected maximum estimation accuracy is given by the Cramer-Rao bound on the squared estimation error, as presented in [13].

4.9 Frequency tracking algorithm

An ENF estimation method based on dynamic programming is presented in [6]. The ENF is estimated by finding the peak locations of the spectrum in each frame. These locations represent the estimated frequency. Dynamic programming is used in order to define a minimum cost path. The magnitude of this cost is calculated by the difference between peak locations in two consecutive frames. This cost function does not allow important frequency jumps between frames and the aforementioned path is employed for the ENF extraction. The ENF is estimated by computing the minimum cost path among the first and the last frame of the signal. The Bayesian Information Criterion (BIC) is used to determine the number of peaks for each frame. Minimizing the BIC criterion yields in the number of peaks n_k. The BIC is formulated as [23]:

$$BIC(n_k) = M\, ln\left(\left\| \mathbf{y} - \sum_{k=1}^{2n_k} \mathbf{a}(\omega_k)\hat{a}(\omega_k) \right\|^2 \right) + 5(2n_q) ln M. \tag{4.15}$$

For frame $k = 1, \ldots, K$, the estimated peak locations which correspond to frequencies are denoted as $P_k = \left\{P_{k1}, P_{k2}, \ldots, P_{qk_k}\right\}$. The core of this method is based on finding a path $\left\{f_k\right\}_{k=1}^{K}$ with peak locations $f_k \in P_k$ and $f_k - f_{k-1} = min$ for $k = 1, \ldots, K$. The set of the estimated peak (frequency) locations, which correspond to the ENF estimation, can be calculated by minimizing the following [6]:

$$C = min_{f_k \in P_k} \sum_{k=2}^{K} (f_k - f_{k-1})^2,\ k = 1, \ldots, K. \tag{4.16}$$

Minimizing the above equation by adopting an exhaustive search is impractical and, thus, a dynamic programming approach in recursively determining the minimum cost path is employed. This is achieved by minimizing the cost from a given frame $j < K$ to the last frame denoted by $J(j, f_j)$:

$$J(j, f_j) = \min_{f_k \in P_k} \sum_{k=j+1}^{K} (f_k - f_{k-1})^2, \ f_j \in P, \ k = j+1, \ldots, K. \tag{4.17}$$

The computational complexity of dynamic programming is of $\mathcal{O}(KP_{max}^2)$, where $Pmax$ is the number of peaks appear in the frame with the highest BIC value. This frequency tracking algorithm based on dynamic programming will be notated as " *method's name (tracking)*" in the thesis.

Chapter 5

Experimental Evaluation

All ENF extraction methods adopted in the thesis are tested on datasets Data 1 and Data 2 using the two choices of frame length L, namely L_1 as in [6] and L_2 shown in Table 3.1. The first choice is used to allow comparisons with the results disclosed in [6], where a novel algorithm for ENF extraction, based on dynamic programming, is suggested. The second one is used for studying the behavior of ENF extraction methods, when longer frames are used. Although longer frame lengths seem to provide higher resolution, their usage may lead to smoother signals, which suffer from the loss of valuable information. In each case, this work suggests that proper filtering and parameter selection of spectral estimation methods can provide high quality ENF estimation.

5.1 Data 1 results

In this section, the results of all methods adopted in this thesis regarding *Data* 1 are presented for both frame length L_1 and L_2. The results are analyzed and comparisons with other works are discussed. Firstly, we proceed by demonstrating the extracted ENF against the reference ENF signal. The extracted ENF signal was obtained by the third harmonic scaled by a factor 3 to fall within the range of ENF. The ENF obtained by STFT is shown in Figure 5.1. The ENF extracted by the Capon method is shown in Figure

5.2. The ENF extracted by the Welch method is depicted in Figure 5.3, while BT and ESPRIT methods applied to ENF extraction are demonstrated in Figure 5.4 and Figure 5.5, respectively. It is worth mentioning that the MUSIC method yields similar results to ESPRIT for both ENF contour and matching accuracy. For illustration purposes all overlaid methods are shifted by $0.05\,Hz$.

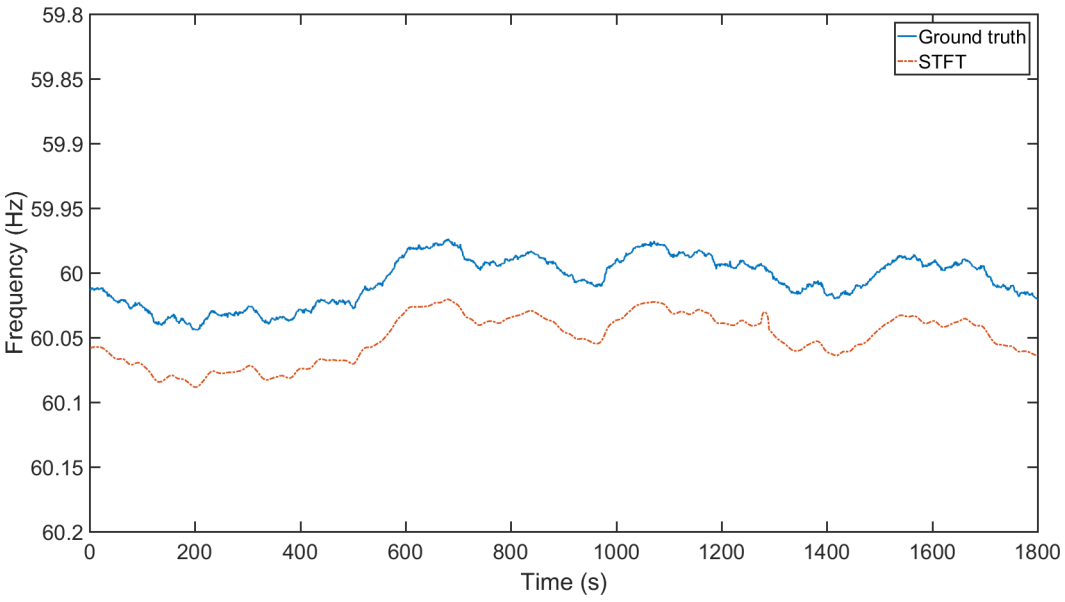

Figure 5.1: Extracted ENF signal against a reference one for the STFT applied to $Data\,1$ for frame length parameter L_1. The extracted ENF signal is shifted by $0.05\,Hz$ for illustration purposes.

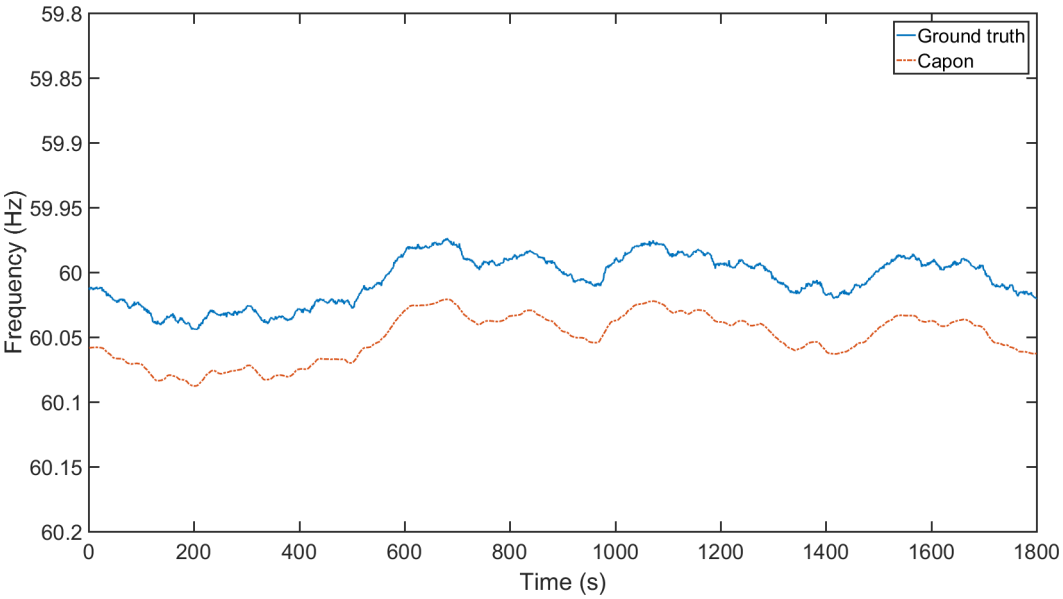

Figure 5.2: Extracted ENF signal against a reference one for the Capon method applied to $Data\,1$ for frame length parameter L_1. The extracted ENF signal is shifted by $0.05\,Hz$ for illustration purposes.

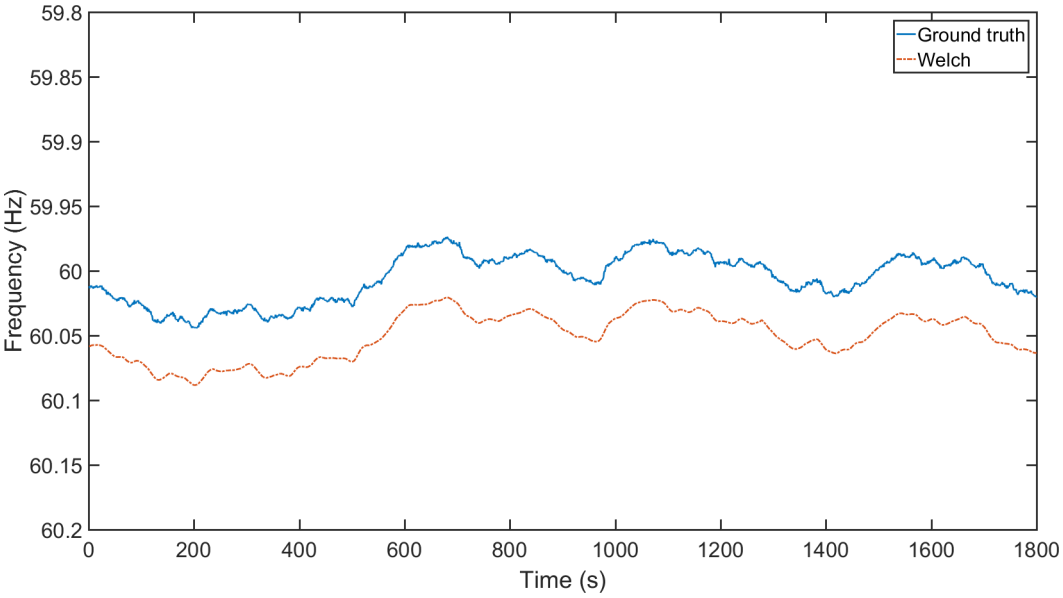

Figure 5.3: Extracted ENF signal against a reference one for the Welch method applied to $Data\,1$ for frame length parameter L_1. The extracted ENF signal is shifted by $0.05\,Hz$ for illustration purposes.

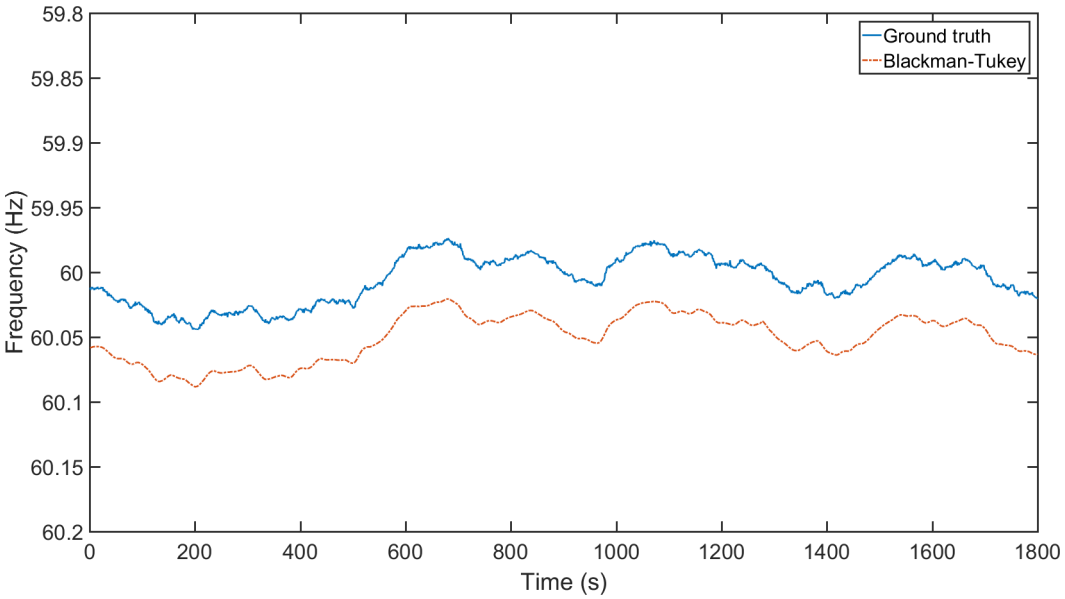

Figure 5.4: Extracted ENF signal against a reference database for the Blackman-Tukey method applied to $Data\,1$ for frame length parameter L_1. The extracted ENF signal is shifted by $0.05\,Hz$ for illustration purposes.

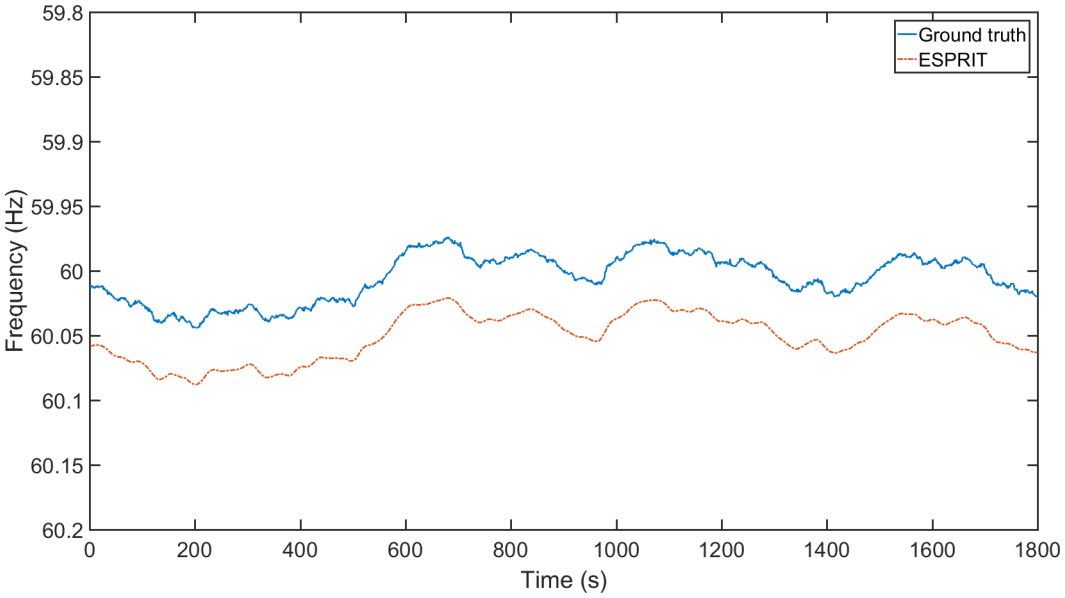

Figure 5.5: Extracted ENF signal against a reference database for the ESPRIT method applied to $Data\,1$ for frame length parameter L_1. The extracted ENF signal is shifted by $0.05\,Hz$ for illustration purposes.

For illustration purposes, the ENF signals extracted from $Data\,1$ by various estimation methods are overlaid in Figure 5.6, when $L = L_1$. The ENF is estimated from the third harmonic down-shifted to $60\,Hz$ scaled by a factor 3. The various ENF contours are shown with increment of $0.05\,Hz$ starting from $59.95\,Hz$. Daniell and BT methods yield similar ENF contours to Welch and STFT, respectively. Similarly, the MUSIC method yields similar results to ESPRIT.

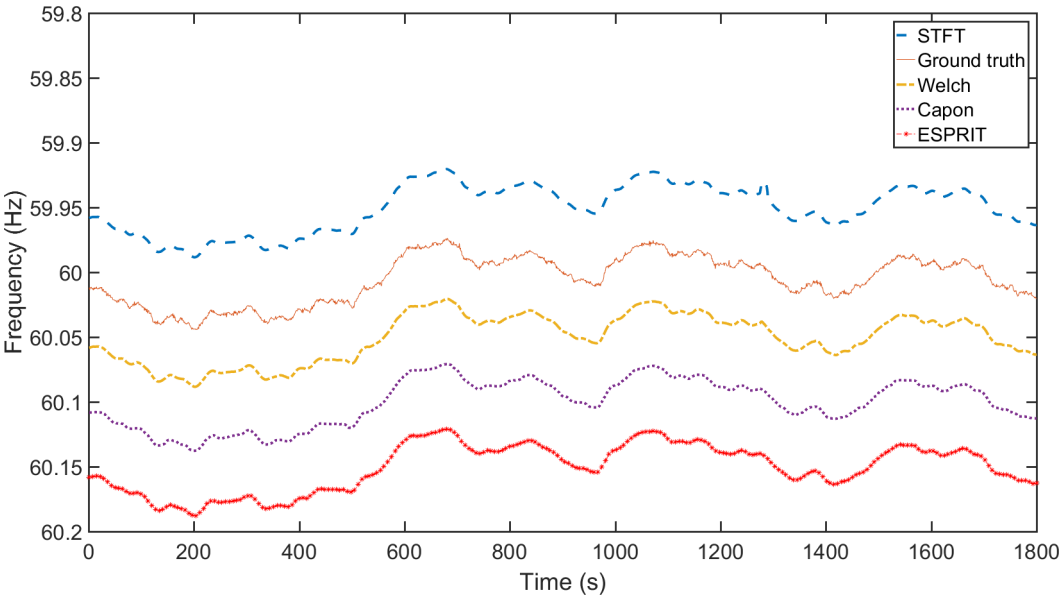

Figure 5.6: Extracted ENF signal of all methods against reference database for $Data\,1$ and frame parameters L_1.

For Data 1 using frame length L_1, the accuracy between the ENF signal extracted by various spectral analysis methods and the ground truth, measured by frequency disturbance recorders with accuracy up to about $\pm 0.5\,mHz$ [6], is summarized in Tables 5.1 and 5.2. The maximum correlation coefficient is listed in Table 5.1, while the minimum standard deviation of error is gathered in Table 5.2. It is seen that the ENF estimated from the first harmonic and the third one is more accurately estimated than that estimated by the second harmonic. For the first and third harmonics, the Welch and the ESPRIT methods yield the best performance with respect to both figures of merit. MUSIC method presented almost identical results to ES-

PRIT for all three harmonics. Although these methods seem to be the most accurate, all algorithms performed very well and this is due to the proper signal filtering and parameter selection. The third harmonic seems to be the most accurate for all algorithms with respect to correlation-coefficient. The minimum standard deviation of errors yields better accuracy in third harmonic than the other two for all methods apart from ESPRIT and MUSIC. The second harmonic is the weaker of the three, but it still provides acceptable ENF estimation. The errors between the extracted ENF and the reference one are at the order of a few mHz.

Table 5.1: Maximum correlation coefficient for various algorithms (Data 1, L_1)

Algorithm	$60\,Hz$	$120\,Hz$	$180\,Hz$
STFT	0.9886	0.985	0.9957
Welch	0.9983	0.985	0.9983
Blackman-Tukey	0.9924	0.985	0.9978
Daniell	0.9906	0.985	0.9977
Capon	0.9952	0.9913	0.9977
ESPRIT	0.9979	0.9913	0.9979
MUSIC	0.9979	0.9913	0.9979

Table 5.2: Minimum standard deviation of error for various algorithms (Data 1, L_1)

Algorithm	$60\,Hz$	$120\,Hz$	$180\,Hz$
STFT	$2.806 \cdot 10^{-3}$	$3.202 \cdot 10^{-3}$	$1.303 \cdot 10^{-3}$
Welch	$1.069 \cdot 10^{-3}$	$3.202 \cdot 10^{-3}$	$1.069 \cdot 10^{-3}$
Blackman-Tukey	$2.284 \cdot 10^{-3}$	$3.202 \cdot 10^{-3}$	$1.218 \cdot 10^{-3}$
Daniell	$2.542 \cdot 10^{-3}$	$3.41 \cdot 10^{-3}$	$1.245 \cdot 10^{-3}$
Capon	$1.806 \cdot 10^{-3}$	$2.9 \cdot 10^{-3}$	$1.237 \cdot 10^{-3}$
ESPRIT	$1.198 \cdot 10^{-3}$	$2.901 \cdot 10^{-3}$	$1.202 \cdot 10^{-3}$
MUSIC	$1.198 \cdot 10^{-3}$	$2.901 \cdot 10^{-3}$	$1.208 \cdot 10^{-3}$

Compared to [6], the accuracy of the ESPRIT method applied to Data 1 is increased from 0.947 to 0.9913 for the second harmonic, which is the weakest. Similarly, the standard deviation of error is decreased from $6.57 \cdot 10^{-3}$ to $2.901 \cdot 10^{-3}$. Also, the most effective method presented in [6] was the

STFT (Tracking), which uses a discrete dynamic programming approach, but still it is not so accurate as the methods presented in the thesis. In Table 5.3 and Table 5.4 the results for STFT (Tracking) are presented for both correlation-coefficient and minimum standard deviation of error. Comparing these findings to the results presented in Tables 5.1 and 5.2 for the third harmonic, we see that the methods developed here, which employ proper filtering and parameter selection, yield a better ENF estimation than that in [6].

Table 5.3: Correlation-Coefficient for STFT (Tracking) (Data 1, L_1)

Algorithm	$180\,Hz$
STFT (Tracking)	99.68

Table 5.4: Minimum standard deviation of error for STFT (Tracking) (Data 1, L_1)

Algorithm	$180\,Hz$
STFT (Tracking)	$1.851 \cdot 10^{-3}$

Another method implemented is the MLE used in [13]. In our work, similar results were found for the MLE algorithm with respect to both standard deviation of errors and correlation-coefficient between the ENF estimated by MLE and reference one, as shown in Table 5.5 and Table 5.6. The standard deviation of errors with respect to reference database for $Data\,1$ was found $0.76\,mHz$ for frame length duration $T = 3\,sec.$ This estimation came from the first harmonic of $Data\,1$, since the two other harmonics (i.e., 2, 3) did not provide better results neither remarkable improvements.. The frame length used in the thesis for MLE is $T = 1\,sec.$

Table 5.5: Correlation-coefficient between the ENF estimated by the MLE and the reference one (Data 1)

Algorithm	$60\,Hz$
Maximum-likelihood estimation	99.91

Table 5.6: Minimum standard deviation of error between ENF estimated by MLE and the reference one (Data 1)

Algorithm	$60\,Hz$
Maximum-likelihood estimation	0.00078

Regarding the implementation of the MLE algorithm is worth noting that increasing the frame length causes a significant drop in accuracy both with respect to correlation-coefficient and standard deviation of error. This is depicted in Figure 5.7 and Figure 5.8. Both figures show the accuracy as the frame length increases till $100\,sec$.

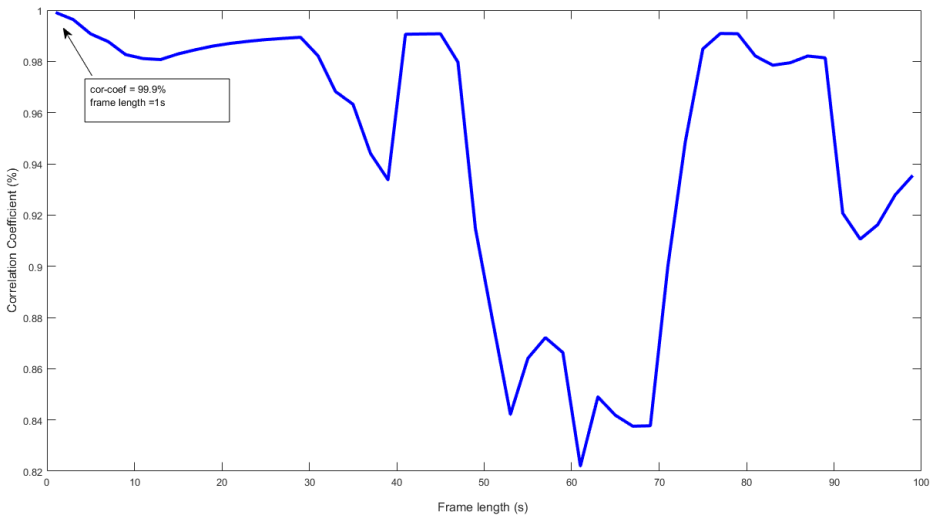

Figure 5.7: Correlation-coefficient for the MLE algorithm versus frame length.

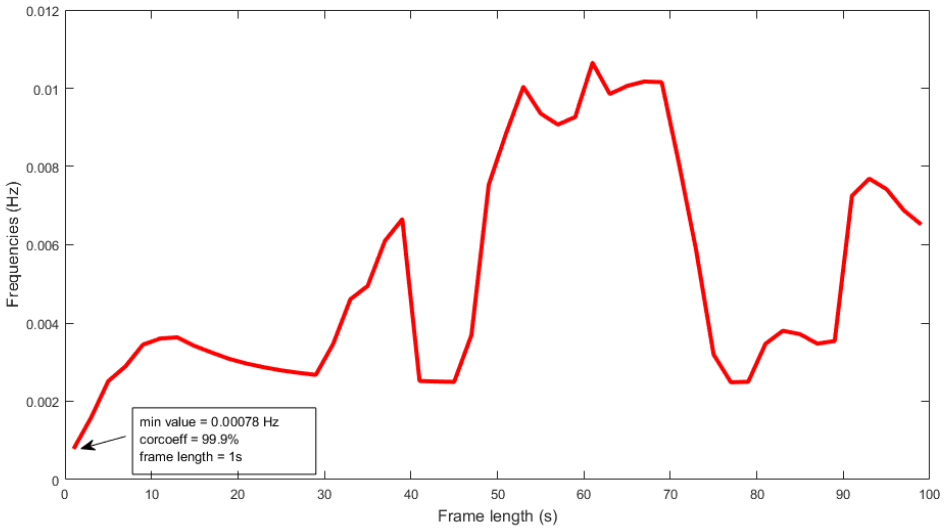

Figure 5.8: Standard deviation of errors for the MLE algorithm versus frame length.

By employing the frame length L_2, which is longer than L_1, one expects an increased SNR and more fine spectral resolution at the cost of lower time resolution. This is evident in Tables 5.7 and 5.8 for Data 1. Frame-by-frame, ENF extraction followed by quadratic interpolation yields more accurate results using STFT with respect to both maximum correlation coefficient and minimum standard deviation of error than using L_1. The first harmonic of ENF is more accurately estimated by the ESPRIT and Welch methods with respect to both aforementioned figures of merit. It is worth noting, that all periodogram-based methods (i.e., STFT, Welch, Daniell, BT) yield a more accurate estimation of the ENF from the second harmonic than the Capon and ESPRIT methods. This is attributed to the "noisy" nature of the estimate of auto-covariance matrix due to the weak SNR associated to this harmonic.

Regarding to L_2 parameter two kind of behaviors are noticed. The first one includes the methods that provide better figures of merit than those when the smaller frame length (L_1) was used. Since a longer frame length is adopted, a better resolution is achieved. The second kind of behavior includes the methods that provide worst figures of merit than those when the

smaller frame length (L_1) was used. It is worth mentioning that the longer frame length makes the extracted signal smoother and valuable information is lost, as shown in Figure 5.9 for the Capon method. An improvement in the extracted ENF is obtained for the STFT method when a longer frame length is used, as depicted in Figure 5.10. Regarding to the second harmonic all methods except the ESPRIT, MUSIC, and Capon perform better when the frame length L_2 is adopted. Using L_2 in the third harmonic, all methods' accuracy drops except the Daniell method. In the latter, the accuracy increases.

Table 5.7: Maximum correlation-coefficient for various algorithms (Data 1, L_2)

Algorithm	$60\,Hz$	$120\,Hz$	$180\,Hz$
STFT	0.9916	0.992	0.9964
Welch	0.9964	0.992	0.9965
Blackman-Tukey	0.9933	0.992	0.9967
Daniell	0.9926	0.9915	0.9967
Capon	0.995	0.9916	0.9953
ESPRIT	0.9953	0.9916	0.9953
MUSIC	0.9953	0.9917	0.9953

Table 5.8: Minimum standard deviation of error for various algorithms (Data 1, L_2)

Algorithm	$60\,Hz$	$120\,Hz$	$180\,Hz$
STFT	$2.388 \cdot 10^{-3}$	$2.317 \cdot 10^{-3}$	$1.572 \cdot 10^{-3}$
Welch	$1.564 \cdot 10^{-3}$	$2.317 \cdot 10^{-3}$	$1.546 \cdot 10^{-3}$
Blackman-Tukey	$2.133 \cdot 10^{-3}$	$2.317 \cdot 10^{-3}$	$1.508 \cdot 10^{-3}$
Daniell	$2.238 \cdot 10^{-3}$	$2.389 \cdot 10^{-3}$	$1.151 \cdot 10^{-3}$
Capon	$1.851 \cdot 10^{-3}$	$3.007 \cdot 10^{-3}$	$1.809 \cdot 10^{-3}$
ESPRIT	$1.793 \cdot 10^{-3}$	$3.007 \cdot 10^{-3}$	$1.804 \cdot 10^{-3}$
MUSIC	$1.793 \cdot 10^{-3}$	$3.007 \cdot 10^{-3}$	$1.795 \cdot 10^{-3}$

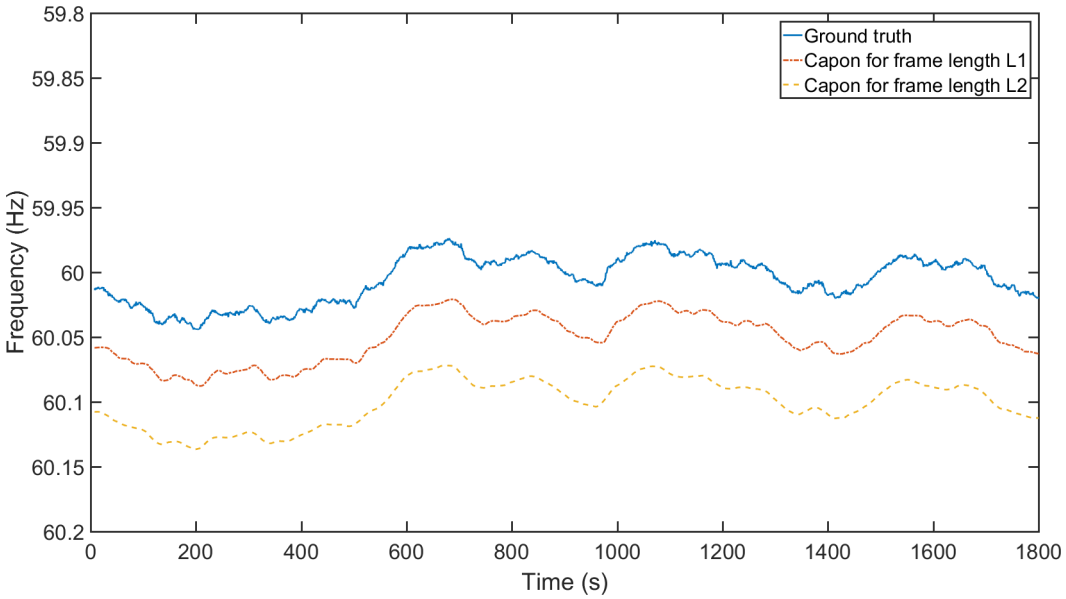

Figure 5.9: Comparison of Capon method for frame lengths L_1 and L_2.

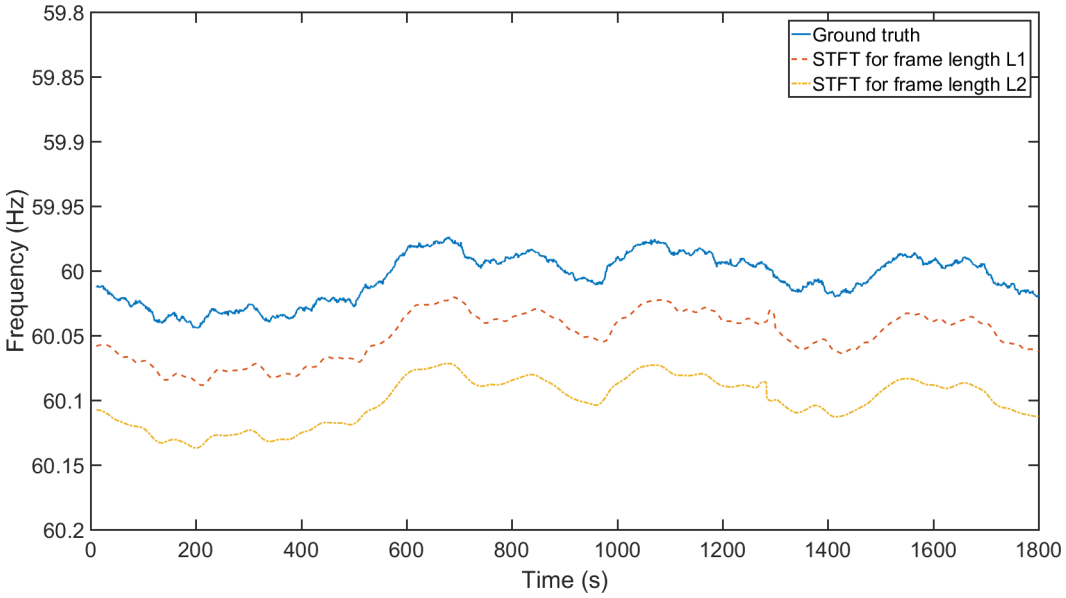

Figure 5.10: Comparison of STFT method for frame lengths L_1 and L_2.

5.2 Data 2 results

In this section, the results of all ENF extraction methods applied to *Data* 2 are presented for both frame lengths, L_1 and L_2. First, we proceed by demonstrating the extracted ENF from the second harmonic against the reference one. The second harmonic of ENF is estimated by various spectral analysis methods. Next, the second harmonic is scaled to $60\,Hz$ in every case. The ENF estimated by the STFT is shown in Figure 5.11. The ENF estimated by the ESPRIT method is shown in Figure 5.12. The ENF estimated by the Capon method is depicted in Figure 5.13. It is worth mentioning that MUSIC yields similar results to ESPRIT for both ENF contour appearance and matching accuracy. For illustration purposes multiple ENF contours are shifted by $0.05\,Hz$ to allow comparisons. As is demonstrated in all figures, more interference is present in the extracted ENF signal from the speech recording than that from *Data* 1.

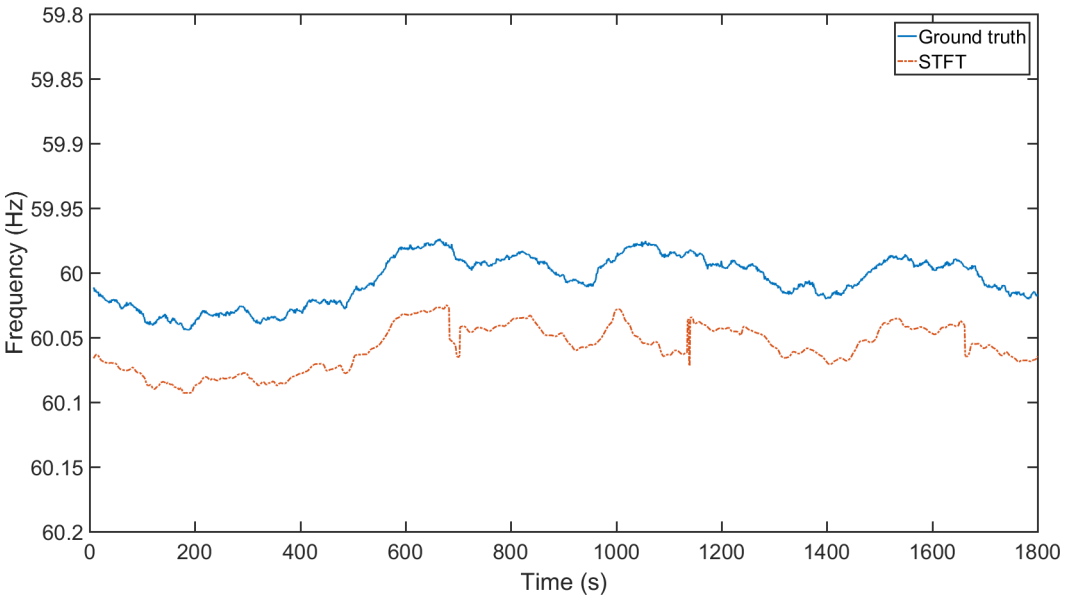

Figure 5.11: Extracted ENF signal against the reference one for the STFT method applied to *Data* 2 for frame length L_1. The extracted signal is shifted by $0.05\,Hz$ for illustration purposes.

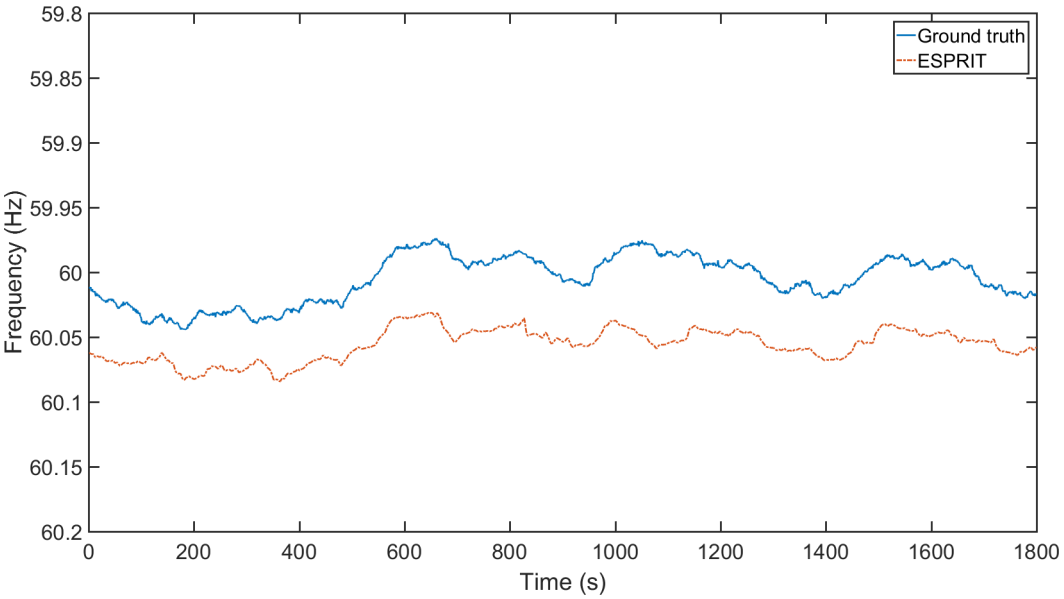

Figure 5.12: Extracted ENF signal against the reference one for the ESPRIT method applied to *Data* 2 for frame length L_1. The extracted signal is shifted by 0.05 Hz for illustration purposes.

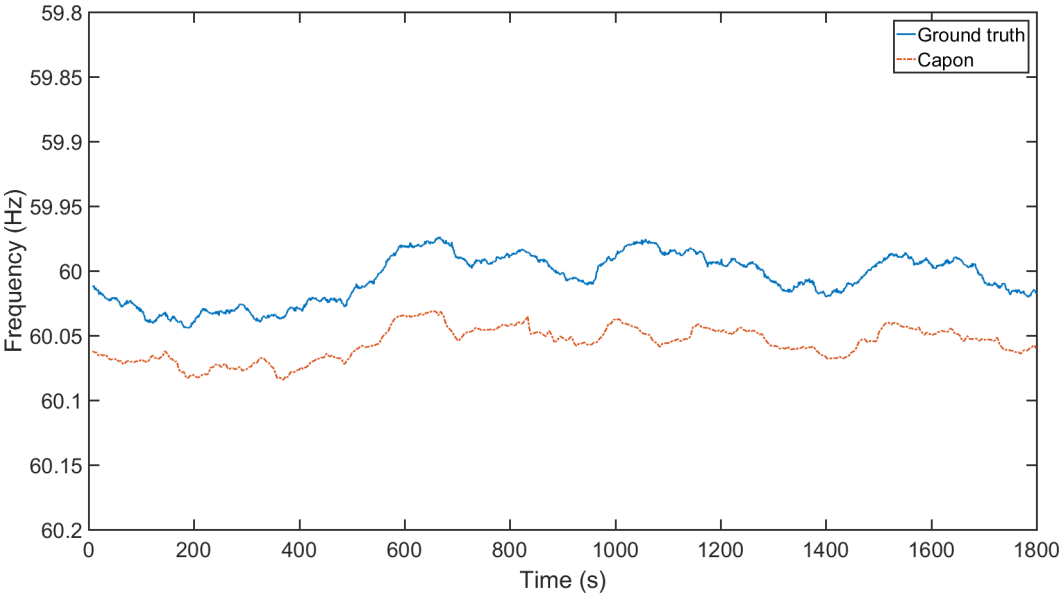

Figure 5.13: Extracted ENF signal against the reference one for the Capon method applied to *Data* 2 for frame length L_1. The extracted signal is shifted by 0.05 Hz for illustration purposes.

For illustration purposes, the ENF contours estimated from the Data 2 second harmonic by various estimation methods are overlaid in Figure 5.14, when $L = L_1$. The second harmonic is down-shifted to $60\,Hz$ and the various ENF contours are shown with increment of $0.05\,Hz$. MUSIC and Welch methods yield similar ENF contours to ESPRIT and Capon, respectively.

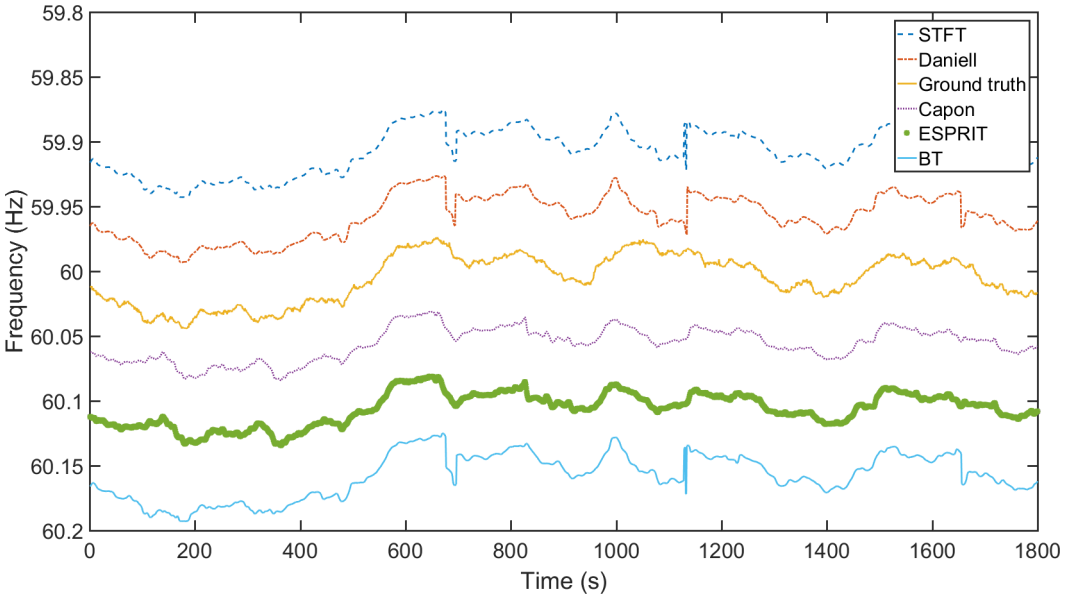

Figure 5.14: Extracted ENF signal of all methods against reference database for *Data* 2 and frame parameters L_1.

Apart from the standard deviation of errors, an interesting insight to the differences between the extracted ENF signal and the reference one is shown in Figure 5.15 and Figure 5.16. Absolute ENF estimation errors from the speech signal are shown to visually demonstrate the performance of the STFT and the ESPRIT.

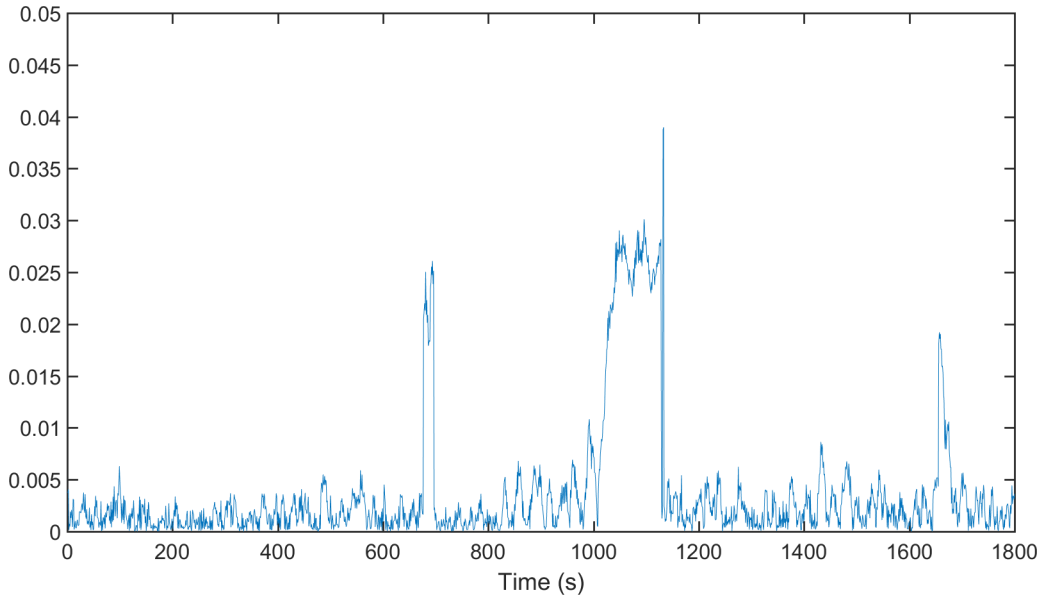

Figure 5.15: Absolute errors between the ENF estimated from *Data* 2 and the reference ENF, when the STFT method with frame length L_1 is applied to the band-pass filtered signal around the second harmonic.

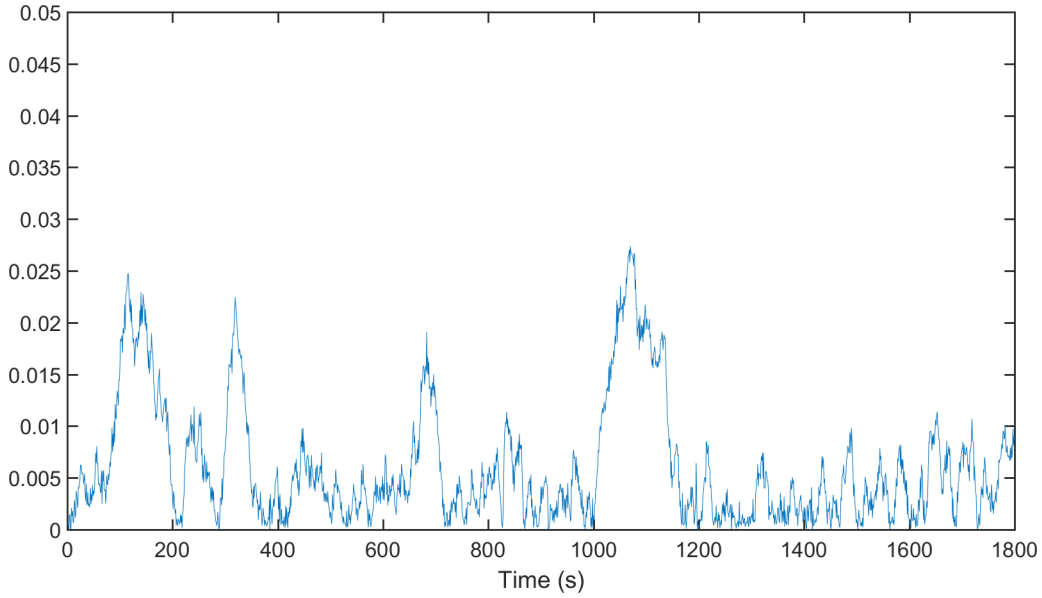

Figure 5.16: Absolute errors between the ENF estimated from *Data* 2 and the reference ENF, when the ESPRIT method with frame length L_1 is applied to the band-pass filtered signal around the second harmonic.

An improvement in the extracted ENF signal is achieved by the ESPRIT method when longer frame length was used, as depicted in Figure 5.17.

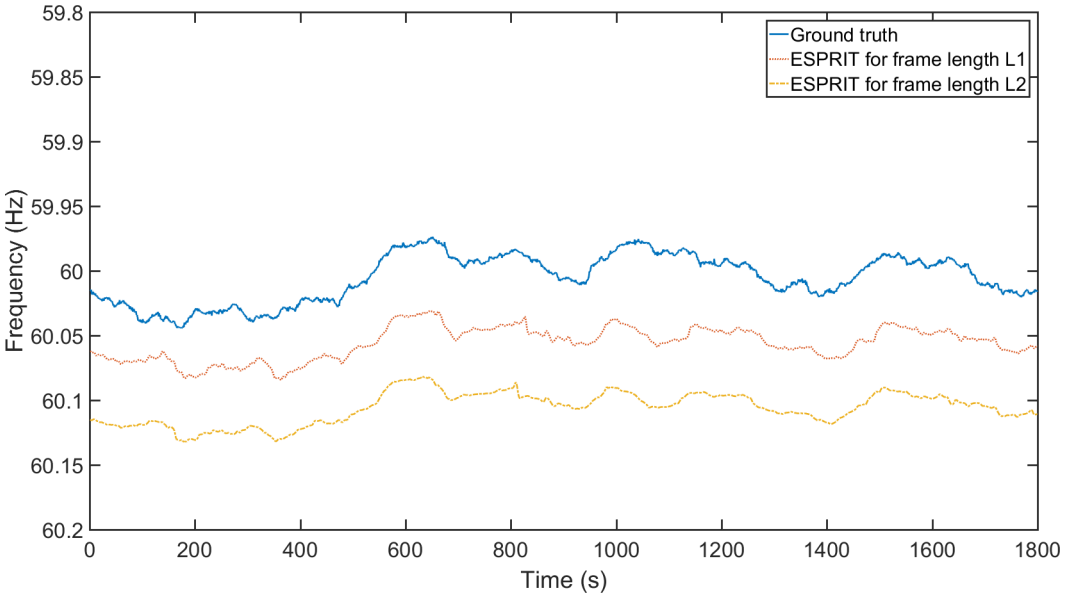

Figure 5.17: Comparison of the ESPRIT method for frame lengths L_1 and L_2 applied to *Data* 2.

In the speech recording, the first and third harmonics of the ENF are too weak [6]. Accordingly, we confine ourselves to the second harmonic ($120\,Hz$). Tables 5.9 and 5.10 summarize the maximum correlation coefficient and the minimum standard deviation of error, respectively. It is seen that the ESPRIT, MUSIC and Capon method yield more accurate results than the periodogram-based methods with respect to maximum correlation coefficient. The maximum correlation coefficient for ESPRIT reported here is 0.9318, while 0.8446 was reported in [6] for the same length L_1. The same ordering of ENF extraction methods is observed when a longer frame length L_2 is employed. Periodogram-based methods yield a maximum correlation coefficient slightly less than the ESPRIT, MUSIC and Capon methods. However, they are ranked top with respect to the minimum standard deviation of error. In Data 2, the application of L_2 frame length yields an improvement in accuracy for all methods with respect to both correlation-coefficient and minimum standard deviation of error.

Table 5.9: Maximum correlation coefficient for various algorithms for Data 2 and two frame lengths L_1 and L_2

Algorithm	$120\,Hz$ (L_1)	$120\,Hz$ (L_2)
STFT	0.9238	0.9332
Welch	0.9179	0.9328
Blackman-Tukey	0.9238	0.9332
Daniell	0.9176	0.9311
Capon	0.9318	0.9444
ESPRIT	0.9318	0.9444
MUSIC	0.9318	0.9444

Table 5.10: Minimum standard deviation of error for various algorithms for Data 2 and two frame lengths L_1 and L_2

Algorithm	$120\,Hz$ (L_1)	$120\,Hz$ (L_2)
STFT	$7.052 \cdot 10^{-3}$	$6.623 \cdot 10^{-3}$
Welch	$7.313 \cdot 10^{-3}$	$6.641 \cdot 10^{-3}$
Blackman-Tukey	$7.052 \cdot 10^{-3}$	$6.623 \cdot 10^{-3}$
Daniell	$7.326 \cdot 10^{-3}$	$6.722 \cdot 10^{-3}$
Capon	$8.011 \cdot 10^{-3}$	$7.797 \cdot 10^{-3}$
ESPRIT	$8.012 \cdot 10^{-3}$	$7.797 \cdot 10^{-3}$
MUSIC	$8.011 \cdot 10^{-3}$	$7.797 \cdot 10^{-3}$

5.3 Computational requirements

The algorithm used in ENF extraction must fulfill two requirements. First and foremost, it has to be efficient in terms of accuracy in order to achieve the best possible matching. The second requirement concerns the computational cost of the method adopted. This is of high importance when an algorithm is implemented. In the thesis, besides the study concerning the evaluation of each algorithm in terms of accurately matching the extracted ENF signal to the reference one, a systematic study is done in order to make clear how each algorithm and raw signal filtering affects the computational requirements. In Table 5.11, the computational time for Data 1 when L_1 is used, is presented. When the frame length L_2 is adopted the computational time is listed in Table 5.12. In both cases, the Capon method is by far the most time consuming with $1394.8667\,sec$ for L_1, while ESPRIT and MUSIC are the next more

time consuming methods. STFT requires the least time (*about* 1.5 *sec*). BT, Daniell and Welch are also very fast.

Increasing frame length to L_2, the order of the most time consuming algorithms remains the same as for L_1. However, all algorithms require more time now. A visual representation of ENF estimation accuracy measured with respect to the correlation-coefficient and computational time requirements of all methods applied to Data 1, using frame length L_2, is shown in Figure 5.18. A logarithmic scale is used on the time axis.

Table 5.11: Computational cost of each algorithm in seconds (Data 1, L_1)

Algorithm	$60\,Hz$	$120\,Hz$	$180\,Hz$
STFT	1.6108	1.7965	1.8208
Welch	13.0709	14.3214	13.2014
Blackman-Tukey	5.4646	5.5289	4.7036
Daniell	2.7730	3.1489	13.2014
Capon	1394.8667	1389.0156	1346.8761
ESPRIT	76.7445	76.4896	76.3824
MUSIC	76.6992	76.5861	76.7358

Table 5.12: Computational cost of each algorithm in seconds (Data 1, L_2)

Algorithm	$60\,Hz$	$120\,Hz$	$180\,Hz$
STFT	2.8796	2.7099	2.7380
Welch	42.7841	43.5214	42.7939
Blackman-Tukey	8.2119	7.5159	7.7323
Daniell	5.8448	5.3647	5.3801
Capon	2752.9680	2807.0183	2770.0832
ESPRIT	157.7661	153.9399	161.4595
MUSIC	158.4205	154.7654	155.3864

For Data 2, there is a significant increase in computational time required by the Welch algorithm, which is even bigger than that of the Capon method. From the inspection of Table 5.13, one sees that for both L_1 and L_2, the Welch method requires a huge amount of time, which makes it completely inefficient for ENF extraction, since it requires approximately 23902 *sec*. This is due to segment length employed in Welch algorithm as stated in Section 4.2. The second most time consuming method is the Capon method, which requires about 2248 *sec*. The rest of the methods seem to be more efficient in terms

of the computational time. The STFT needs the least time (about 2.96 sec) along with the Daniell and BT methods. When the frame length L_2 is used, the time is slightly increased in all cases, as expected. Except the Welch and the Capon methods, the rest fulfill the requirements for fast and accurate ENF extraction.

Table 5.13: Computational cost of each algorithm in seconds (Data 2, L_1 and L_2)

Algorithm	$120\,Hz\,(L_1)$	$120\,Hz\,(L_2)$
STFT	2.9648	3.2584
Welch	23902.8156	27320.0221
Blackman-Tukey	6.4617	9.3799
Daniell	4.8143	7.0967
Capon	2248.7507	3424.2689
ESPRIT	126.2766	208.2021
MUSIC	125.1108	198.7045

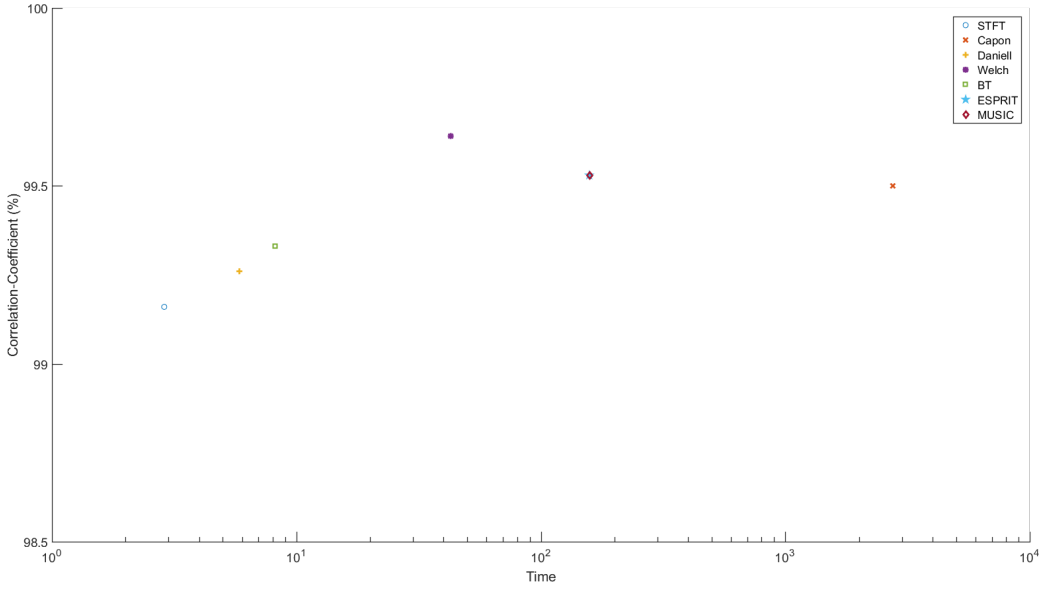

Figure 5.18: Correlation-coefficient for the first harmonic ENF extraction versus computational requirements of the various mathods applied to *Data* 1 using frame length L_2.

Chapter 6

Conclusion

Digital audio authentication requires high accuracy in ENF extraction to ensure exact time/location estimation. Several frequency estimation methods have been tested on a frame-based approach by dividing the entire signal into consecutive overlapping frames. Two datasets have been employed, namely, a signal recorded from the power mains and a speech recording. In the former dataset, a high SNR is present and accurate ENF estimation can be achieved from all the three harmonics. In the latter dataset, strong interferences are present. It has been demonstrated by experiments that if the raw datasets are filtered by a properly designed band-pass filter, then either non-parametric or parametric techniques for spectral estimation provide an accurate estimation of the ENF. The aforementioned interferences introduce challenges to ENF estimation. Exploiting their sparse nature in the formulation of ENF estimation could be a topic of future research. By applying longer frames it has been shown that one can reduce the interferences and achieve an accurate ENF. This longer frame setup fits better the speech dataset, where strong interferences are known to exist. The computational time requirements are also studied for both datasets, because computational speed is of high importance as well. As expected, when longer frames are employed, computational time increases. This is also noticed, when a higher filter order is used for band-pass filtering the raw signal. When computational time is considered, the STFT and Daniell are found to be the fastest methods, offering both

accuracy and low computational time. From the point of view of computational time, the Welch and the Capon methods are found to be the most time consuming ones.

Bibliography

[1] R. C. Maher. Audio forensic examination. *IEEE Signal Processing Magazine*, 26(2):84–94, 2009.

[2] C. Grigoras. Digital audio recording analysis: the electric network frequency criterion. *Int. Journal Speech, Language, and the Law*, 12(1):63–76, 2005.

[3] C. Grigoras. Applications of enf criterion in forensic audio, video, computer and telecommunication analysis. *Forensic Science International*, 167 2-3:136–45, 2007.

[4] C. Grigoras. Applications of ENF analysis in forensic authentication of digital audio and video recordings. *Journal of the Audio Engineering Society*, 57(9):643–661, 2009.

[5] A. Hajj-Ahmad, R. Garg, and M. Wu. Instantaneous frequency estimation and localization for ENF signals. In *Proc. 2012 Asia-Pasific Signal and Information Processing Association Annual Summit and Conf.*, pages 1–10, 2012.

[6] O. Ojowu, J. Karlsson, J. Li, and Y. Liu. ENF extraction from digital recordings using adaptive techniques and frequency tracking. *IEEE Trans. Information Forensics and Security*, 7(4):1330–1338, 2012.

[7] D. P. Nicolalde-Rodriguez, J. A. Apolinario, and L. W. P. Biscainho. Audio authenticity based on the discontinuity of ENF higher harmonics. In *Proc. 21st European Signal Processing Conf.*, pages 1–5, 2013.

[8] A. J. Cooper. The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings. an automated approach. In *Proc. 33rd AES Int. Conf. Audio Forensics-Theory and Practice*, 2008.

[9] A. J. Cooper. An automated approach to the electric network frequency (ENF) criterion: Theory and practice. *Int. Journal Speech, Language, and the Law*, 16(2):193–218, 2009.

[10] R. Garg, A. Hajj-Ahmad, and M. Wu. Geo-location estimation from Electrical Network Frequency signals. In *Proc. 2013 IEEE Int. Conf. Audio, Speech, and Signal Processing*, pages 2862–2866, 2013.

[11] M. Huijbregtse and Z. Geradts. Using the ENF criterion for determining the time of recording of short digital audio recordings. In *Proc. Int. Workshop Computational Forensics*, pages 116–124, 2009.

[12] R. Garg, A. L. Varna, and M. Wu. Modeling and analysis of electric network frequency signal for timestamp verification. In *Proc. 2012 IEEE Int. Workshop Information Forensics and Security*, pages 67–72, 2012.

[13] D. Bykhovsky and A. Cohen. Electrical network frequency (ENF) maximum-likelihood estimation via a multitone harmonic model. *IEEE Trans. on Information Forensics and Security*, 8(5):744–753, 2013.

[14] Y. Hu, C.T. Li, Z. Lv, and B. Liu. Audio forgery detection based on max offsets for cross correlation between ENF and reference signal. In *Proc. The Int. Workshop on Digital Forensics and Watermarking 2012*, pages 253–266. Springer, 2013.

[15] A. Hajj-Ahmad, R. Garg, and M. Wu. Spectrum combining for ENF signal estimation. *IEEE Signal Processing Letters*, 20(9):885–888, 2013.

[16] R. Garg, A. L. Varna, and M. Wu. Seeing ENF: Natural time stamp for digital video via optical sensing and signal processing. In *Proc. 19th ACM Int. Conf. Multimedia*, pages 23–32, 2011.

[17] R. Garg, A. L. Varna, A. Hajj-Ahmad, and M. Wu. Seeing ENF: Power-signature-based timestamp for digital multimedia via optical sensing and signal processing. *IEEE Trans. on Information Forensics and Security*, 8(9):1417–1432, 2013.

[18] H. Su, A. Hajj-Ahmad, R. Garg, and M. Wu. Exploiting rolling shutter for ENF signal extraction from video. In *Proc. 2014 IEEE Int. Conf. Image Processing*, pages 5367–5371, 2014.

[19] J. O. Smith and X. Serra. *PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation*. CCRMA, Department of Music, Stanford University, 1987.

[20] J. B. Allen and L. R. Rabiner. A unified approach to short-time fourier analysis and synthesis. *Proc. of the IEEE*, 65(11):1558–1564, 1977.

[21] P. Stoica and R. L. Moses. *Spectral Analysis of Signals*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.

[22] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. on Antennas and Propagation*, 34(3):276–280, 1986.

[23] P. Stoica, J. Li, and H. He. Spectral analysis of nonuniformly sampled data: A new approach versus the periodogram. *IEEE Trans. on Signal Processing*, 57(3):843–858, 2009.