UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Forensic audio: The relevance of Electric Network Frequency Analysis as an authentication tool in South Africa

by

Tarryn-Chevonne Morgaine Denton

A mini-dissertation submitted in partial fulfillment of the requirements for the degree

Master of Music (MMus) (Music Technology)

Department of Music

Faculty of Humanities

University of Pretoria

Supervisor: WM de Villiers

2014

ABSTRACT

The predominance of digital media in modern society has presented a need for methods of authenticating digital audio. When an audio recording is presented as part of legal or business proceedings, it may be necessary to provide proof of authenticity. This mini-dissertation investigates to what extent Electric Network Frequency (ENF) analysis can be implemented as an authentication technique within a South African context.

The electricity supply frequency exhibits random deviations in time from the nominal frequency of 50 Hz. ENF analysis is based on the premise that the supply frequency is detectable in audio recordings made on equipment utilising, or in close proximity to, electrical equipment that utilises the main electrical network. The recorded ENF signal can be extracted and used as a unique temporal identifier for determining the authenticity of an audio recording.

Results of experiments conducted in this study indicate that it is possible for a detectable ENF signal to be induced into an audio recording, and subsequently isolated for further analysis. It is possible to compare the extracted ENF signal to the extracted ENF signal from a reliable database to determine temporal authenticity, and investigate the integrity of the file. It is found that the following minimum requirements must be met for successful ENF authenticity analysis:

- The recording device be powered by the main electrical network,
- The equipment be proven to be suitable for use in ENF analysis, i.e. no UPS device is used; the microphone does not have any electrical components that could alter the ENF signal, and the sound card does not induce any frequency components or noise into the signal; and
- The availability of a reliable comparative temporal database of ENF values.

Furthermore, results indicate that establishing an ENF temporal database by recording the supply frequency utilising a data acquisition unit, and subsequently processing the signal, is more practical than recording the database as an audio file. It was found that the South African electricity supply is sufficiently random to be suitable for ENF authenticity analysis. Recommendations for further study are made.

KEYWORDS OF THE STUDY

- Electric Network Frequency (ENF)
- Digital Audio Authenticity
- Forensic Audio
- Audio Integrity
- Authentication
- Digital Audio Recording

ACKNOWLEDGEMENTS

I would like to express my gratitude and heartfelt thanks to the following people:

- My supervisor, Murray de Villiers, for the years of support, encouragement and guidance;

- My friend and colleague, Rudi Kroch, for his advice and help with the processing of the audio signals; and

- My loved ones for their unending support and encouragement during the course of this study.

LIST OF FIGURES

CONTENTS

TERMINOLOGY AND CONCEPT CLARIFICATION

The following definitions are adapted from a multitude of different prominent sources within the field of digital signal processing. Among the most referenced sources are:

- Bob Katz – Mastering Audio (2007)
- David Huber & Robert Runstein – Modern Recording Techniques (2005)
- Roey Izhaki – Mixing Audio (2008)
- Mackie – Glossary of Pro Audio Terms (ND) [accessed 4 March, 2013]

The reader is referred to the following list of terms, abbreviations and acronyms used in this study:

Alternating Current (AC) – An electrical current in which the direction of the flow of the electrons reverses at regular recurring intervals or cycles resulting in a voltage that alternates in polarity over time.

Artefacts – An error in the perception or representation of information introduced by the applied equipment or techniques.

Audio Editing – The process of the intentional manipulation of recorded audio data and its associated metadata in order to alter, adapt, or refine to achieve a particular purpose or result.

Audio Enhancement – The process of audio editing that results in the improvement of audio perception such as speech intelligibility and audibility of low intensity level sounds within a recording.

Audio Authentication – Confirming authenticity of a recording in terms of being a true, unabridged reflection of the events or representation of the facts. In legal terms, an authentic document is a document that is exactly what it claims to be.

Audio Recording – To register sound in a reproducible permanent form by mechanical, digital or electrical means. For the purposes of this study, the term audio recording refers to any form of recorded audio including, but not limited to:

- Recorded conversations,
- Recorded statements,

- Recorded interviews,
- Recorded board meetings,
- Audio feed from video such as camera or CCTV footage, and
- Studio and live music recordings.

Audio Transcriptions – The conversion of recorded material from one storage format to another. For example:

  - Vinyl to CD transcription
  - Analogue tape to CD transcription
  - Tape/CD to text transcription

Bootleg Recordings – An audio or video recording of a performance that was not officially released by the artist or under any other authorised representative authority. Bootleg recordings include recordings of live performances and material created in private or professional recording sessions.

Digital Aliasing – This refers to the creation of supplementary complimentary frequencies as a result of intermodulation distortion.

Digital Audio Workstation (DAW) – A dedicated recording and editing software application and hardware system that generally offers capabilities such as:

- Sequenced multi-track recording, editing and mixdown,
- Audio signal processing, and
- Integration of supplementary peripheral hardware devices and software interfaces

Digital Signal Processing (DSP) – The manipulation of digital audio data and metadata by means of the application of a transfer function in the digital domain. The goal of DSP is usually to produce measurement, spectral, temporal, dynamic and spatial manipulations of the audio data.

Direct Current (DC) – An electrical charge that flows consistently in one direction. Batteries (with a definite positive and negative terminal) produce a direct current with constant voltage polarity.

Dropout – An audio signal defect caused by a temporary loss of signal.

Downsampling – In signal processing, downsampling is the process of reducing the sampling rate and bit depth of a signal, resulting in a concomitant reduction in the data rate.

Electric Network Frequency (ENF) – The time varying frequency of the electricity supplied by electricity facilities.

Fast Fourier Transform (FFT) – A Fourier Transform is a mathematical technique that can be utilised to convert time domain functions and data to frequency domain functions and data. A fast Fourier Transform is a shortened version of the Fourier Transform algorithm.

Hanning Window – A function that is applied to a data set before it is subjected to Fourier analysis. The application of the window function improves frequency resolution accuracy, although usually at a loss of time domain amplitude accuracy.

Metadata – Data that provides information about one or more aspects of the data, such as:

- how the data was created, manipulated and applied
- time and date of creation
- author of the data

Native File Format – The file format that is addressable by the default central processing hardware of a specific system, which could include the combination of computer software applications or computer hardware. In the context of software applications, it refers to the file format that the application works with when producing, editing, or publishing the file.

Real Time Processing – When the evaluation of the output data of a system is temporally synchronous (or near synchronous) to the processing of that data. As opposed to offline processing, real time processing produces and evaluates the results as the data is being processed.

Sampling – DSP takes periodic samples of a changing audio waveform and, through a quantization process, transforms the signal levels into a representative stream of binary words. This data can then be manipulated and stored for processing and reproduction.

Sampling Rate – The number of samples (measurements) per second that are used to digitally quantize an analogue audio signal.

Signal-to-noise Ratio (SNR) – A numerical value (often expressed in decibels) that compares the amplitude of the considered signal relative to the average amplitude of the background noise floor.

Short-Time Fourier Transform (STFT) – An application of the Fourier Transform algorithm typically used to provide information about amplitude and frequency variation of a signal over time.

Uninterruptible Power Supply (UPS) – An electrical device that provides battery backup power to a system when the input power source (typically the electrical network supply) fails. A UPS device can also protect equipment from power fluctuations and surges that may damage equipment.

Waveform – A graphic representation of an audio signal in terms of positive and negative attributes being applied to the compression and rarefaction of the conductive medium (air) by the longitudinal wave progression of sound propagation. Characteristics such as frequency and amplitude can be deduced from the resulting representation of compression as a function of time.

Zero-crossing Measurement – The point in a waveform where the polarity changes from positive to negative, represented by a crossing of the time-axis (zero value) in the graph of the function. The number of zero-crosses is used to estimate the fundamental frequency in speech processing.

CHAPTER 1

INTRODUCTION

Multimedia is a ubiquitous aspect of modern everyday life in the civilised world. Opinions and judgements are based upon the content of this multimedia presented to society in the form of audio recordings, news websites, camera footage, and phone conversations (Gupta, Cho & Kuo, 2012:50). There have been significant changes to the types of criminal activities that could generate digital evidence as a result of the proliferation of computers and electronic devices since 1970. The extent of the development in multimedia technology has resulted in the possibility of almost every crime having digital media associated with it (NCR, 2009:179-180). This rapid technological development in the media industry in conjunction with the increasing availability of digital signal processing equipment and advanced audio manipulation software, has led to a significant increase in the number of attempts to utilise digital audio as evidence by both public and legal authorities (Grigoras, 2009:643). Unfortunately, these advances in digital signal processing have also increased the potential for manipulating recorded digital audio material. Audio evidence is therefore increasingly vulnerable to modification and alteration after original production (Cooper, 2006:1).

Questions relating to the integrity of a document or record may arise in matters relating to business, finance, civil, and criminal investigations (NRC, 2009:163). When the standardised procedures for evidence handling fail to ensure authenticity, the reliability of the evidence may depend upon technical analysis to prove the integrity and originality of the recording. Procedurally, the party that wishes to include the audio recording as evidence is responsible for proving the authenticity of the recording. They must demonstrate that the recorded audio has not been altered, edited, or technologically compromised in any way since the time of its production (Cooper, 2006:1). Methodologies that verify authenticity can aid examiners and render the results of analyses more reliable (SWGDE, 2010:2). This has resulted in a need for standardised procedures and protocols when dealing with digital audio material as evidence.

1.1 Background to the study

Audio recording is a commonplace practice in modern society for the capture and archiving of the proceedings of events. There are increasing incidences within the legal and corporate arena that involve the use of audio recordings either as evidence or as part of legal proceedings (Kajstura, Trawinska & Hebenstreit, 2004:165). In order for the recording to be admissible as legal evidence, it has to meet established legal proof criteria. The field of audio forensics provides the methodologies, techniques, and practice guidelines that make it possible to submit audio recordings as evidence within a judicial setting.

Audio forensics can be applied when:

- Audio recordings are utilised as evidence in a judicial setting,
- When referring to a recorded interview or meeting where verbal agreements were made,
- Unauthorised recordings and audio-visual piracy are investigated, and
- Archives are recorded and extracted.

The forensic analysis of digital audio media usually occurs when (Koenig & Lacey, 2009:663):

- There are unexplained sounds or dropouts in the recording,
- There are specific technical questions relating to discrepancies between written statements and recorded audio statements, and
- Most frequently; to determine if any editing was performed, whether it is a duplicate, and whether or not the recorder was stopped and started again at any time during the audio recording.

Audio forensic services are utilised by law enforcement authorities; professionals in the legal field such as investigators or lawyers; commercial agencies, such as banks and other financial institutions; and companies within the music sector that would benefit from reliable archival methodologies, such as the South African Music Rights Organisation (SAMRO), the South African Recording Rights Association Limited (SARRAL), National Organisation for Reproduction Rights in Music (NORM), and the Recording Industry of South Africa (RiSA).

There are multiple aspects of audio forensics, such as voice identification and elimination, audio enhancement for purposes of transcriptions, crime scene and timeline recreation, and authentication (Maher, 2009:84). This study focuses on authentication; specifically the use of Electric Network Frequency (ENF) analysis as a temporal authentication technique.

1.2 Statement of the problem

Recent technological advances provide for the general availability of sophisticated multimedia devices, some of which can be valuable in providing evidence or supporting information for use in legal proceedings (Watney, 2009:4).

Digital audio files are highly susceptible to manipulation, alteration and falsification (Koenig & Lacey, 2009:683), which can be difficult to detect. This implies a need for methods of authentication on digital media (Maher, 2009:87). To qualify for use in the legal or business sector, these authentication techniques need to be reliable, assured, objective and repeatable.

The present work investigates ENF analysis as an authentication method for digital audio recordings. Some of the proposed scenarios that would warrant the application of ENF analysis within the context of audio technology are:

- To determine originality of recordings;
- Validation of the date and time of recorded events for archival purposes; and
- To determine the integrity and authenticity of evidence, including the detection of evidence tampering in litigation or legal proceedings.

1.3 Applicability of ENF analysis

In order to investigate the applicability of ENF analysis as a tool for authentication, the following questions are posed:

- Does South Africa have sufficient infrastructure in place to use ENF analysis as an authentication tool?
- Is the electricity supplied to the end user suitable for ENF analysis, i.e. free of artefacts, in order to be extracted from audio recordings?

- Can ENF analysis be used in a South African context for the purposes of authenticating a digital audio recording?

Many modern societies interact in a predominantly digital medium. The digital medium includes the use of devices such as computers, mobile phones, printers, iPods, tablets, and digital cameras for the generation and storage of information as well as for communication (Watney, 2009:2). There is a vast amount of digital information generated by these devices, some of which take the form of audio and video recordings of conversations, meetings and events within the corporate and entertainment sectors. The amount of information as well as the ease of use of editing software has resulted in the need for reliable methods of authentication. Authentication can provide the examiner with information such as the start and stop time of a recording, the location and time of a recording, and the presence of any irregularities within the recording. In order for digital audio/video material to be used in official proceedings as evidence, it has to be proven to be authentic and contain no alterations. Similar studies done elsewhere have concluded that ENF analysis is a reliable method of authentication for use on digital audio recordings (Grigoras, 2003: 4). This study investigates the applicability of this technique on recordings done in a South African context.

1.4 Aim of the study

There is little research available on the topic of ENF analysis in South Africa. This study investigates the theory of ENF analysis for application within a South African context, exploring the field of validity for the application of this analysis type in South Africa. This is achieved by conducting a critical review of available literature, followed by a practical validation of the theory by means of empirical experimentation.

The aim of this study is to provide legal practitioners and audio technologists with an introductory background into ENF authentication methodology and principles. This study is of value to the audio technologist, persons with audio technology experience, companies, or persons engaging the legal/forensics sector through the field of audio forensics and ENF analysis.

Research goals include:

- Verification of the likelihood that a questioned recording has recorded the ENF signal on equipment powered by the electrical network or (DC) battery powered equipment.
- Validation of ENF analysis as a method of temporal authentication or manipulation detection on digital audio recordings.
- Establishing whether the ENF signal can be extracted and used for comparison to a known database for temporal authentication.
- Ascertaining if the electricity supply in South Africa is suitable for use in ENF temporal authentication analysis.
- Investigate the possibility of establishing a reliable reference database for ENF temporal authentication analysis.

The aim of this study therefore is to validate ENF analysis theory as a method for forensic authentication, and to provide an introduction for audio forensic practitioners in South Africa.

1.5 Research questions

Given that authentication of digital audio recordings is a concern, the aim of this study is to ascertain the applicability of ENF analysis to determine the temporal authenticity and integrity of a digital audio recording. This poses the following research question:

- To what degree can ENF analysis be used as an authentication tool in South Africa?

In support of the research goals, the following sub-questions should be addressed:

- What is ENF analysis and its applicability within audio forensics?
- Under what conditions can the ENF be extracted from a digital audio recording and utilised for the purposes of authentication?
- What are the limitations of ENF analysis?
- What infrastructure is required for the application of ENF analysis to audio forensics? Including:
    - What is the role of the reference database?
    - To what extent can ENF analysis be done without the reference database?
    - If the analysis can be done, what are the limitations of the results?

1.6 Delimitations of the study

The field of audio forensics is a wide area of study that encompasses many different sectors. This study focuses on the authentication of digital audio recordings from an audio technical point of view. It is assumed that the reader of this study will have some background knowledge of digital signal processing techniques.

The field of study is limited to:

- Authentication of digital audio recordings by means of ENF analysis; and
- Temporal authentication and manipulation detection of digital audio recordings.

The study specifically excludes:

- Analogue recordings;
- Telephone conversations recorded on a mobile telecommunications device;
- Validation of ENF analysis relevant equipment;
- Audio forensic methodologies other than authentication;
- Advanced digital audio processing techniques used in audio production;
- Judicial evidentiary proceedings and chain of custody procedures; and
- Evidence collection.

Automated decision algorithms as a technique of ENF analysis is not addressed, and the topic of phase discontinuity only briefly discussed. No emphasis is placed on evaluating the differences in the various ENF signal search methods.

The compression algorithms of some recording devices can affect the susceptibility of that device to record the ENF disturbance. In order to simplify the evaluation, and not extend past the scope of this study, only uncompressed audio files are used for evaluation.

The fundamental concepts of electricity supply within the network grids have been studied in depth by numerous international authors (Grigoras, 2009; Michelek, 2009; Kajstura, Trawinska, and Hebenstreit, 2005; Kantardjiev, 2011; Chang and Huang, 2011). The above authors have found that the deviations in the frequency are random in nature. It must be noted that the qualities that define randomness are not addressed within the scope of this study, and is suggested as a topic of further research. Due to the similar properties of the electricity supply in South Africa to that of Europe, it is assumed the entire network will carry the same frequency. As such, no further research is conducted on the frequency supply value at different points within the network grid. For the purposes of ENF analysis and audio forensics, the parameter analysed is limited to supply frequency.

This study does not address recordings produced by uninterrupted power supply devices, and private stand-alone generators, only recordings produced on equipment powered by the national electrical network or dry cell batteries are analysed.

CHAPTER 2

LITERATURE REVIEW

Communication in society functions in primarily physical and electronic media; the latter being utilised in information and telecommunication devices such as computers, mobile phones, printers, iPods, iPads, tablets, and digital cameras to name but a few. These advanced electronic devices have led to increased opportunities to commit crimes such as fraud, and evidence tampering (Watney, 2009:2-3). Digital signal processing techniques could be used to conduct piracy over the Internet, manipulate evidence intended for legal use, and modify security device recordings (Gupta, Cho & Kuo, 2012:50). Likewise, parties involved in legal proceedings may depend upon information generated, stored or distributed on electronic devices. Electronic information may be relevant in proving or disproving a fact or point in question relating to the guilt or innocence of an individual and/or company. This constitutes electronic evidence and the rules of the law of evidence are applicable thereto in deciding the admissibility of the information as evidence (Watney, 2009:2-3).

2.1 Audio forensics

Digital devices and personal computers are ubiquitous in today's modern society. It is standard practice for telephone conversations to be recorded by institutions and companies such as banks and financial service providers, and to capture events on digital devices such as cell phone cameras. The cost of acquiring this technology is relatively inexpensive, and the operational requirements are often basic. In addition, there is a multitude of easily accessible training material available on the Internet in the form of training manuals, instructional videos and courses on editing digital media. Operational requirements, such as extended recording time, smaller size, and portability, have resulted in devices that are suitable for implementation in the corporate and legal sector, but at the cost of low bit-rates, reduced quantization, and harsh compression algorithms (Koenig, Lacey & Killion, 2007:352-353). As a result, most evidentiary recordings have many undesirable qualities, such as limited frequency response, restricted dynamic range, interfering environmental noise and distortion. Under these conditions, the transition from analogue to digital technology has not necessarily resulted in better quality audio recordings. It does, however, allow for additional processing opportunities to aid the forensic examination (Koenig, Lacey & Killion, 2007:353).

Audio evidence refers to any audio material that may be used as evidence in legal proceedings (Cooper, 2006:1). Audio evidence is typically obtained as part of an investigation or as part of an inquiry into an accident or civil incident (Maher, 2009:84). These recordings can originate from investigations by legal or law enforcement professionals as well as recordings supplied by witnesses, victims, and suspects (Cooper, 2006:9).

Editing recordings made in the "real world" (in comparison to audio recording studio conditions) can be fairly complex due to the nature of the surrounding environment. Submitted recordings usually contain high level background noise such as voices, music, television and radio broadcasts, traffic noise and other sounds; limited signal to noise ratios; increased distortion; compression artefacts; and effects of poorly placed microphones. These factors make it fairly difficult to edit a recording by combining or chaining small fragments without the sequence sounding unnatural (Koenig & Lacey, 2009:669). However, these effects can also result in a recording that is less than ideal for use as evidence. Recordings that contain the above mentioned qualities would benefit from the application of audio forensics in order to enhance the quality of the audio to improve speech intelligibility, remove interfering background noises, as well as authenticate the recording.

Three of the main sub-sections of audio forensics are (Maher, 2009:84):

- Enhancement of audio evidence in order to improve speech intelligibility and the audibility of low level sounds;
- Audio forensic procedures, which includes: interpreting and documenting sonic evidence, such as identifying talkers, transcribing dialogue, and reconstructing crime or accident scenes and timelines; and
- Establishing the authenticity of audio evidence.

2.1.1 Audio enhancement

Forensic audio typically contains elements that impede proper analysis such as noise, distortion and other interfering sounds. The goals of forensic enhancement are either to improve intelligibility or reduce listener fatigue for the purposes of speech transcription, or to help reveal idiosyncratic background sounds that may provide important information (Maher, 2009:88). Recordings that would benefit from the application of enhancement contain audio signal that has been degraded prior to, or during the recording of the event. The application of digital signal processing in these instances requires a high degree of specialised skill, experience and equipment. It is necessary for the examiner to have knowledge of time and frequency domain signal processing, acoustics, psychoacoustics, and speech production (Cooper, 2006:9). This process is usually performed using an exact copy of the original so that the original remains unaltered.

An initial critical listening examination of the media will determine if the audio requires enhancement and, if so, to what degree. The principle enhancement procedures include time-domain level detection and frequency domain filtering (Maher, 2009:88).

Time domain level detection refers to processing that is done on the amplitude of the recorded audio signal (Maher, 2009:89). It is possible to control the effectiveness of time domain methods by making use of advanced digital signal processing techniques, and manipulation of the filter's various parameters (Maher, 2009:90).

Frequency domain filtering methodologies generally utilise spectral subtraction. This involves calculating an estimate of the noise level as a function of frequency and subsequently subtracting the result from the input signal (which contains the unwanted noise). The enhanced output is created by reconstructing the signal from the subtracted spectrum. As with time domain enhancement, this technique can be improved by incorporating digital signal processing techniques to reduce the effects of unwanted artefacts (Maher, 2009:90).

2.1.2 Audio forensic procedures

This involves the documentation and interpretation of audio such as identifying talkers, transcribing dialogue, and reconstructing crime or accident scenes and timelines (Maher, 2009:84). Within this section of audio forensic analysis, one would typically find techniques, such as gunshot analysis, voice identification, and signal processing for the purposes of transcription, amongst others. It is not feasible to compile a complete list of these techniques, since the number of potential signal analysis procedures is too large (Koenig & Lacey, 2009:683). As such, only two of the main techniques are briefly discussed, namely documentation and reporting, and voice identification/elimination.

Documentation and reporting involves a detailed report describing and interpreting the evidence, the methods used, and the statistical basis for any conclusions reached (Maher, 2009:91). The expert must prepare detailed work notes relating to the major aspects of the examination. These notes usually contain comprehensive details about the submitted recording, i.e. all devices and software used in the examination (including the relevant settings thereof), all graphs and charts produced as a result of the analysis, results of the examinations, and any other observations or comments regarding the audio (Koenig & Lacey, 2009:685).

All findings should be objectively listed in the report. For example, if the examination revealed no evidence of an alteration to the media, but the file format is such that it renders it impossible to scientifically detect, the examiner should note this in the report and include a detailed list of recommended methods, equipment, conditions and expertise needed to further analyse the file (Koenig & Lacey, 2009:685).

Voice identification/elimination within the context of audio forensics is essentially the process of deciding whether one voice is the same as another (or not), i.e. whether each voice segment belongs to the same speaker (Gupta, Cho & Kuo, 2012:57). If the samples cannot be matched as the same speaker with certainty, the tested voice sample can be eliminated. In contrast, voice identification aims to identify the talker (Rumsey, 2008:211).

2.1.3 Authentication of digital audio recordings

The most common requests for forensic analysis are those concerning the integrity of the digital audio file (Cooper, 2006:10). The transition from analogue to digital media has complicated the question of authenticity of digital audio. Evidence of tampering or editing on a digital file is not as obvious as with analogue magnetic tape (Maher, 2009:87). Manipulations can be impossible to detect in a reliable way by aural evaluation alone. It is fairly easy to alter or manipulate a digital audio file without requiring access to specialised equipment and training (Rappaport, 2012:1).

Digital audio refers to the reproduced sound that is heard through a digitised representation of discrete samples (Rappaport, 2012:13). Digital media can include (Rappaport, 2012:1):

- Portable digital audio recorders,
- Voicemail machines or servers,
- Compact Discs (CDs) or Digital Versatile Discs (DVDs),
- Digital Audio Tapes (DATs) or minidisc recorders,
- Hard drives,
- Flash memory drives,
- Mobile phones,
- Surveillance equipment,
- Digital videos with audio tracks, etc.

Authentication examinations are conducted to determine (Koenig & Lacey, 2009:669):

- The originality of the recorded information,
- If the file has been altered in any way (relates to the integrity of the audio file),
- If the file contains any discontinuities (such as recorder stops or starts), and
- Whether or not it matches a specific recording system, if presented for analysis.

The Audio Engineering Society (AES) defines an authentic recording as a recording that is made simultaneously with the acoustic events it claims to record, and using the same methods and equipment as claimed by the party responsible for the production of the recording; a recording free from any unexplained artefacts, alterations, additions, deletions or edits (Grigoras, 2009:643). In his thesis, Rappaport suggests the necessity to separate the word accurate from the definition of authentic. This results in 2 categories of authentic recordings: A) authentic and accurate, and B) authentic and inaccurate. Category B would constitute “unaltered recordings that contain discontinuities introduced in real-time without human intention, such as voice operated recorder (VOR) pauses, dropouts, packet losses, Radio Frequency (RF) interference as well as unaltered recordings containing mechanical stops/starts or hoaxes (reciting dialogue) that are authentic, but is not a complete or accurate portrayal of events” (Rappaport, 2012:19). Rappaport suggests the division of the question of authenticity into two parts, allocating the terms legal and technical authenticity to each section. Legal authenticity relates to the need for proof that the material is original and is what it claims to be (an accurate representation of the events). Technical authenticity relates to claims regarding the likelihood that the characteristics of a recording match those of an original recording (the authenticity of an audio recording) (Rappaport, 2012:17). A further challenge lies in differentiating between intentional and unintentional modification (Rumsey, 2008:212). A discontinuity could be the result of a number of different variables, such as (Rappaport, 2012:37):

- Signal interruption (e.g. a cable disconnection);
- Transmission dropouts;
- Radio frequency interference;
- Disc write or buffer overrun errors;
- Manual pauses;
- Automatic voice operated recording;
- Internal editing performed on the device; and
- Post-production processing performed on a Digital Audio Workstation (DAW).

The above list is not complete and only provides some examples of causes of discontinuities. It is important to bear these in mind when conducting an authentication analysis. Many of the above-mentioned variables are only applicable to some recording methods and devices and should therefore be considered, or excluded, as possible causes of discontinuities.

In their 2009 paper, Koenig and Lacey suggest that it is normally not possible to definitively authenticate converted copies of original recordings, since it is difficult to detect some alterations after a recording has been copied onto a computer system, changed or converted into another format, and subjected to compression and recompression. There are many indicators of originality that can be influenced by the process of copying and converting the file. Amongst these is the consistency between the non-audio metadata present in the submitted recording and in test recordings made on the alleged original recorder (or similar model). In addition to this, the duplication process may remove some of the features that could aid in the authentication examination, such as:

- Artefacts and dropouts consistent with editing;
- Unnatural speech interruptions;
- Phase discontinuities; and
- Discrete frequency tones present in the original recording.

It is often not possible to be 100% certain of the originality of a recording, but an analyst can state that the characteristics of a recording are consistent with those of an original submitted file (Rumsey, 2008:212). Although none of these techniques are reliable enough to scientifically analyse a recording for the purposes of audio forensics, they can help to generate a wide range of supportive evidence that can be used in a court or legal institution (Rumsey, 2008:217).

2.2 Requirements for admissibility of digital audio as evidence

As a forensic science discipline, the analysis of digital evidence refers to the gathering, processing and interpretation of digital and multimedia data. The following digital devices which can generate and store digital evidence that could prove valuable in criminal investigations include (NRC, 2009:179):

- Desktop and laptop computers,
- Cell phones,
- Digital cameras,
- Personal digital assistants,
- Large servers and storage devices, and
- Portable media players (e.g. iPods).

The storage media associated with the above mentioned devices can be categorised into three broad categories:

- Magnetic memory (hard drives and tapes),
- Optical memory (CDs and DVDs), and
- Electronic storage (Universal Serial Bus (USB) flash drives, memory cards and some microchips).

In *Admissibility of electronic evidence in criminal proceedings: An outline of the South African legal position,* Watney (2009) describes electronic evidence as any information that is generated, distributed or stored on an electronic device that may be useful in proving or disproving a fact or point in question relating to the guilt or innocence of the accused. A digital audio recording is assumed to be authentic, and allowed as evidence, unless a party within the proceedings questions the authenticity of the recording. Regarding the use of digital documents or data messages as evidence, the South African law states that the party wishing to use the data as evidence has to satisfy the court of the documents' authenticity. An expert will be required to testify to the authenticity of the recording through the process of a trial within a trial. Requirements for the admissibility of digital evidence are regulated by the Electronic Communications Act. This act defines a data message (digital evidence) as data generated, sent, received or stored by electronic means and includes voice and stored records. In digital form, audio is considered to be susceptible to error and falsification (Watney, 2009:8).

For evidence to be deemed admissible in a court of law or other official venue, the techniques used in order to carry out the investigation must be:

- Proven to the relevant authorities to be unbiased and objective,
- Statistically consistent (reliable) and repeatable,
- Non-destructive, and
- Recognised by experts within the field.

Techniques used may not result in a change of the original meaning or understanding of the evidentiary recording (Maher, 2009:84-85). According to the Forensic Speech and Audio Analysis Working Group (FSAAWG, 2009:4), the following minimum requirements must be met in order to validate the ENF techniques and procedures with regard to the quality of results:

a) The critical aspects of the analysis have been identified and the limitations defined.
b) The methods, equipment and software used have been demonstrated to be fit for use in ENF analysis.
c) The ENF analysis is fully documented.
d) The results obtained from the analysis are reliable and fully reproducible, and limitations are clearly indicated in the analysis notes and/or final report.
e) The ENF technique/procedure has been subjected to independent assessment and where new, peer review.
f) The technical specialist conducting the analysis has demonstrated that they are competent to do so.

Software testing processes should involve the identification of the key purposes of ENF analysis and have a documented test procedure to ensure that it fully meets the requirements proposed for the validation. It is imperative to record the experimental setup and different parameter values of the applied software tools in order to be able to reproduce the analysis process. Factors of the analysis such as sampling frequency, FFT size, window overlap span, window size and type etc. must be carefully detailed. Data backups of the analysis results are essential.

2.3 Digital audio authentication techniques

Digital audio authentication is the complex process by which the origin of a questioned recording is investigated in order to determine its consistency with an original recording (if available), or to investigate the possibility of manipulation (Rappaport, 2012:15). Confirming authenticity proves that the recorded events are genuine and that the record can be relied upon (Cooper, 2006:12). Audio authentication is generally understood to be every kind of documentation that can verify the origin and the contents of the audio recording and recording media in question (Rappaport, 2012:14). The purpose of the authenticity examination is to ascertain whether the recording is original and to provide possible explanations for any identified anomalies in the recording, such as discontinuities in the recorded signal. In addition, the examination can provide any extra information or reinforce some information claimed about the recording, recording system, and the environment in which the recording was made (Cooper, 2006:10).

In his thesis on the detection of copies of digital audio recordings, Cooper suggests splitting the question of authenticity into two parts that relate to the authenticity and accuracy of the recording respectively. In the part relating to the authenticity of the recorded audio, the examiner aims to establish the consistency between the actual recorded event and what is known or alleged about the incidents in question. For example: that the voices in the recording are consistent with the people present at the time of the recording, establishing that the date and time when the recording was produced are true, and confirming that the location where the audio was recorded is consistent with the digital media. Determining the accuracy of a recording relates to establishing the originality of the recorded digital audio. By confirming the originality of the recording, the examiner eliminates the possibility of the occurrence of various forms of manipulation. When examining the accuracy of a recording, the main concerns of the examination are to determine:

- If the recording is an original or a copy,
- Whether or not the recording has been edited or modified since its production,
- If the recording was made on the recording system it was claimed to have been made, and
- Whether or not the recording is consistent with what is known about the actual recorded events.

An authentic recording is defined as a recording that is made simultaneously with the event it claims to have recorded, and in the same way and method that the party responsible for the production of the recording claims. It is a recording that is free from unexplained artefacts, alterations, deletions or edits. An accurate recording is either an original or an exact copy of an original recording of events that is consistent with the limitations of the recording methodologies, and is fully, exactly and completely faithful to the original events. This relates to the volume and frequency characteristics of the signal over time, as well as the sequence and duration of the events (Rappaport, 2012:17).

Techniques that use the signal and its properties for authentication are regarded as passive techniques. On the contrary, methods that use additional information that is added to the signal for the purposes of authentication (such as watermarking) are regarded as active techniques. Stable tones that last for the duration of the recording are considered best for use with passive authentication methods (Gupta, Cho, & Kuo, 2012:51). When devising a methodology, the examiner must always be aware of what question needs to be answered, i.e. the purpose of the examination. This will determine what techniques should be used, in what order they should be applied, as well as the interpretation of the results (Rappaport, 2012:8). The examiner must investigate the capabilities and limitations of the device used to produce the recording in order to determine an appropriate method of examination. Once familiar with the device, the examiner should be able to explain the reasons for any anomalies in the file that are caused by the recording system. Tools that compare characteristics such as content, data, and time windows within a questioned file, to either itself, exemplar files, or a database, are dependant on the known characteristics of the file format in order to produce statistically reliable results. It is important to be aware of how the file format produces, handles, organizes and stores data (Rappaport, 2012:15-16). In his thesis on establishing a standard for digital audio authenticity, Rappaport suggests that it is necessary to differentiate between container authenticity and content authenticity. The container and the content should be considered as two separate units. Container analysis refers to the file name, file structure and the metadata of the file. Content analysis refers to the bits and bytes that make up the audio portion of the file and the event it reproduces. It is possible to inadvertently affect the authenticity of the container of the media by something as simple as renaming the file, but the integrity of the audio contents will not be compromised. The altered container may raise doubts about the authenticity of the contents of the file. There are several possibilities to consider when dealing with container and content authenticity:

1. Authentic container and authentic content.
2. Authentic container and inauthentic content.
3. Inauthentic container and authentic content.
4. Inauthentic container and inauthentic content.

In these cases, the third option could be further divided into intentional and unintentional modification. Unintentional modification would refer to a mislabelled container. Merely opening and saving a file may change the metadata and cause values to differ from previous analyses. It is necessary to state in the examination report whether the authenticity examination was conducted on the media as a whole package or only on the contents (which is considered the fundamental issue being examined) (Rappaport, 2012:9-10).

The following is a list of equipment suggested by Koenig and Lacey that is required for the purposes of forensic authentication examinations (Koenig & Lacey, 2009:664):

- Professional stand alone digital playback and recording systems
- Proprietary digital playback systems
- Digital data imaging and analysis software
- High-resolution waveform analysis software
- Fast Fourier transform analysers and software
- Spectrographic analysis software
- Professional headphones
- High-speed computers, external sound cards, and high resolution monitors and printers
- Professional audio editing and playback software
- Professional grade amplifiers, cables, connectors etc.
- Equipment racks.

During the course of the authentication examination, the examiner should aim to use a variety of techniques and methods in order to arrive at statistically reliable results. Methods of authentication can be categorised into basic and advanced methodologies. Critical listening, waveform analysis, and frequency spectrum analysis are regarded as basic tests that can detect obvious changes or unusual characteristics or sections that warrant the application of more specialised techniques (e.g. digital data analysis, phase continuity analysis, and residual ENF analysis) (Gupta, Cho, & Kuo, 2012:52). The following sub-chapters contain brief descriptions of some of the most commonly used methods for authentication.

2.3.1 Critical listening

Critical listening is one of the earliest stages of the authenticity examination and involves an overall aural examination of the medium with high-end reference headphones in order to determine the general characteristics of the recorded information. This includes the setting in which the recording was made, overall quality, and factors limiting the intelligibility of the recording. The purpose of critical listening is, most notably, to ensure optimal playback; note the intelligibility of the recording; observe acoustic events (where the recorder was placed, number of talkers, etc); and study the presence of both background and foreground noises (Rappaport, 2012:31). This analysis provides a good point of departure for the continuing authenticity examination, since it could identify any apparent alterations or irregularities, as well as any audible discontinuities such as edits, splices, changes in background sounds, buzzes and tones, etc that require further analysis (Maher, 2009:87).

Aural cues, that the examiner should listen for, include (Cooper, 2006:16):

- Variation in the background noise,
- Transient-like sounds of any nature,
- Audio dropouts,
- Changes in the quality of the audio,
- Unnatural speech or voice characteristics,
- Acoustic quality of the recording environment,
- Artefacts produced by the recording system, e.g. transients, and
- Hums, tones or any form of electronically generated noise.

In their paper *Forensic analysis of digital audio recordings,* Koenig and Lacey divide the critical listening review into four different, but overlapping, areas of interest (2009:673):

- Preliminary overview,
- Possible record and edit events,
- Background sounds, and
- Foreground information.

The preliminary overview is an important first step since it often identifies the portions that require further evaluation and analysis. It allows for the detection of obvious record events, and generally structures and facilitates additional analysis. The remaining tests are normally repeated several times, not in any particular order, both before and after a specific analysis.

Examples of events requiring further investigation include:

- Sudden, unnatural, or uncharacteristic changes in an individual's voice or cadence regardless of the spoken language;
- Inconsistency between the beginning and end times given by the operator and the actual length of the recording;
- Abrupt and unexpected changes in the topic of conversation; and
- Digital compression artefacts and bandwidth characteristics that are inconsistent with the recorder allegedly used to record the conversation.

In order to accurately identify the record and edit events, it is important to critically evaluate the same functions done on test recordings produced on the submitted recording system. The examiner must have good knowledge of editing functions and the results of digital signal processing on an audio file using both consumer and professional grade software. Experience in examining record events is also essential as this analysis involves the close scrutiny of the transient-like signatures left by some digital recorders and editing processes during stopping, starting, adding or deleting and combining sections of files together (Koenig & Lacey, 2009:673). Professional grade headphones are always used for this type of analysis as the playback environment affects some qualities of the recording and can result in less than optimal playback. Headphones also result in better isolation from background noises and reverberation, and the accuracy of the amplitude and frequency response is not dependent on the position of the examiners head. There is also a stable stereo image and near-identical input into both ears (Koenig, Lacey & Killion, 2007:362).

Although the critical listening evaluation is helpful in identifying areas of interest and for determining the order and techniques used in the examination, the results of the analysis are not always objective or quantifiable. The critical listening evaluation should not be relied upon too heavily to reach scientific conclusions (Rappaport, 2012:35).

2.3.2 High-resolution waveform analysis

High-resolution waveform displays are graphic representations of the time and amplitude functions of an audio signal. The horizontal x-axis represents the time function and the vertical y-axis represents the amplitude function (Rappaport 2012:35). These high-resolution waveforms allow for the identification and detailed analysis of events such as (Koenig & Lacey, 2009:674):

- Dropouts and transient-like sounds,
- DC offset,
- Clips and peaks,
- Areas containing zero digital amplitude, and
- Possible edits or discontinuities.

This method of analysis is often used both before and after the critical listening analysis in order to examine recorder stop/start events, or any other points of interest. The strength in this technique lies in its ability to accurately investigate the temporal and amplitude characteristics of a digital audio file. The amount of detail in the waveform display itself can be the size of an individual sample point or the expanse of the entire digital file (Koenig & Lacey, 2009:674). However, it is impractical to examine the entire range of the audio file and signal at a high resolution in search of possible edit points as this could result in the impractical analysis of possibly millions of samples. It is therefore recommended to either use an automated approach and/or critical listening analysis to identify possible points of interest that require further analysis by means of high-resolution waveform analysis (Rappaport, 2012:37). A recording leader section may provide potentially valuable information regarding the recording system such as system noise, DC samples, or non-speech information. This information could be useful in different authentication techniques.

Within the context of waveform analysis, a suspected edit could present as a slight change in the noise floor between spoken words (Koenig & Lacey, 2009:674). Rappaport (2012:35) suggests that a waveform with a drastic change in amplitude, that is not typical of sounds that occur naturally, may be an indication of an edit. In this regard, it is important to analyse the recording device and eliminate any sounds that are caused by the recording device.

2.3.3 Spectrographic analysis

Spectrographic analysis is a 3-dimensional display of the time vs. frequency vs. amplitude characteristics of an audio signal (Koenig & Lacey, 2009:676). A spectrogram can reveal some irregularities in the frequency content, for example manipulations of an audio file consisting of audio sections that were recorded in different environments and subsequently spliced together. It is also useful for the visual examination of stable tones that last for the duration of the audio recording (Gupta, Cho, & Kuo, 2012:52). This form of analysis is used to identify recording system characteristics like the frequency response of the system, and to identify discrete tones within the recorded material that are specific to a recording system. These discrete tones, e.g. the induced mains power line frequency, may be valuable in proving that a recording was produced on a particular system, as well as ascertaining the authenticity and integrity of the digital media (Cooper, 2006:17).

Apart from revealing information regarding the recording system characteristics, spectrographic analysis can be useful in identifying the precise pitches of particular sounds or speech; hence revealing the effects of aliasing and channel/transmission characteristics, certain record events, editing, and varying discrete frequency components (Rappaport, 2012:39). Spectrographic analysis is considered a passive audio authentication technique since it is based on analysing the properties extracted from the signal itself (in this case the frequency spectra contributed by the recording environment) (Gupta, Cho, & Kuo, 2012:52). Although spectrographic analysis can clearly illustrate certain types of events and editing, it cannot be used as a substitute for high-resolution waveform and narrowband spectrum analyses. This type of analysis could also produce misleading or even false examination results (Koenig & Lacey, 2009:677).

There are numerous user-controlled parameters available in analysis software that allows the examiner to optimise the spectrographic output for the specific questioned portion of the recording. This enables the examiner to optimise the display by manipulating user-defined parameters in order to get the best resolution to examine a questioned sound or event. Typical elements that would be examined by means of narrowband spectrum analysis are (Koenig & Lacey, 2009:647):

- Digital aliasing;

- Sample frequency bandwidth;
- Questioned signals;
- Background sounds;
- Convolutional and transmission characteristics; and
- Power-line frequency components.

Analyses of questioned signals, background sounds, and convolutional and transmission characteristics are conducted by comparing the frequency spectrum of the questioned sound or event to that of known sounds and system features (Koenig & Lacey, 2009:675). Convolutional characteristics are best identified over long term spectral averaging, which minimises the effects of vocal and other dynamic signals (Koenig & Lacey, 2009:675).

If the frequency characteristics are time variant or are random in nature (such as the signal induced by the ENF), it is recommended to use a waterfall spectrum display (Koenig & Lacey, 2009:675). A waterfall display is a 3-dimensional spectrogram of the sound with the time function on the x-axis, frequency on the y-axis, and amplitude as indicated by colour intensity or grayscale range (Rappaport, 2012:39). Rappaport (2012:39) suggests the use of spectrum analysis to determine the presence or absence of the ENF signal.

2.3.4 Digital data analysis

When audio information is digitised, the process converts the acoustic information into binary information. This information is stored in a logical structure of bits and bytes, and then interpreted by the DAW (Rappaport, 2012:16). Digital data analysis involves the use of specialised software to analyse the binary information relating to the audio file. The software renders a visual representation of the actual bits and bytes that comprise both the audio and administrative data information (referred to as metadata) (Koenig & Lacey, 2009:678). The examiner typically views and analyses the data in hexadecimal format (Koenig & Lacey, 2009:679). This analysis must be conducted using the digital data image, files exported from the digital data image, or an exact copy that completely and accurately captures the native file format. Examinations of converted files that do not retain the native file format (e.g. PCM wavefiles or MP3 files) usually do not provide information regarding the digital data content of the submitted recording. These conversions normally modify the format of the digital audio and exclude the metadata altogether (Koenig & Lacey, 2009:678).

While edits to the audio may not be audible to the examiner, the digital data could show evidence of tampering, transcoding or resaving. Many of these changes are permanently embedded into the metadata (Rappaport, 2012:16). The metadata of an audio file may contain information regarding the titles of any external software that may have been used on the file, as well as any post-production operations (Hong & Yin, 2013:2927). The metadata is considered very fragile and can inadvertently be changed by simple operations (such as turning the device on, or opening a file on a computer that does not have a write protection). Interactions with external software may result in changes in the metadata that may render the file unplayable. While the metadata may change, the audio itself will likely not change (Rappaport, 2012:28).

Digital data analysis is useful for the purposes of device authentication. By making multiple test recordings using the same recorder, but different settings, it is possible to verify how the unit records time stamps within the metadata. This assists the examiner in identifying any unique characteristics in the way the recorder structures its metadata (Rappaport, 2012:29). Depending on the device used, a header section may be present and contain information regarding the model, serial number, firmware version, time, date and length of the recording according to the internal clock settings. The time stamp contained in the metadata should not be considered faultless since a user can manipulate the internal clock setting on a device. However, this could provide a good starting point to search an ENF database for external temporal authentication (should an ENF signal be present). Test recordings done on the original recording device or one of the same make and model may be important to compare the resulting header information to ensure consistency (Rappaport, 2012:31-32).

2.3.5 Phase continuity analysis

Audio phase refers to the pattern of peaks and troughs in a sinusoidal waveform. The phase of the audio is the measurement between the relative locations of the peaks and valleys in the wave (Rappaport, 2012:40). A waveform analyser is used to create a visual representation of the audio phase. The phase of a waveform is considered to be continuous if the peaks and valleys are consistent between the relative points of measurement. An abrupt shift in the pattern of the phase in either direction is considered a phase discontinuity (Rappaport, 2012:40).

The ENF signal is the most commonly used discrete signal for the purposes of phase continuity analysis (Koenig & Lacey, 2009:684). The premise in using a stable tone at a particular pitched frequency is that, if the recording has been edited or manipulated, it is possible that the phase would not align accurately at the location of editing. Therefore, the edit would appear as a discontinuity in the phase of the signal (Chang & Huang, 2011:12). High precision phase analysis involves the direct comparison of a discrete signal in the submitted recording with a corresponding reference signal. A stand-alone phase meter or waveform analyser is used to display the value of the angular differences between the two signals over time, or to produce a direct statistical comparison between them (Koenig & Lacey, 2009:684).

The applicability of phase analysis can be limited by poor signal-to-noise ratio and the instability of the discrete signal within the submitted recording (Koenig & Lacey, 2009:684).

2.3.6 Authentication software

These software tools typically provide for faster than real-time analysis of submitted recordings. Some of the different analyses that are possible with software applications are (Koenig & Lacey, 2009:684):

- Phase continuity analysis;
- Detection of changes in the background noise characteristics;
- Tracking of discrete signals throughout a recording; and
- Detection of prior digitisation or aliasing artefacts, etc.

The use of software analysis tools may provide for a more convenient and efficient method of analysis; however, the results of the automated analysis should not be accepted without manual verification by the examiner. The results must be very thoroughly scrutinised by the examiner since they are prone to numerous false results (Koenig & Lacey, 2009:684).

Software programs can be divided into two categories, i.e. manual and automated. In the context of ENF analysis, manual analysis software allows a technical specialist to check for the presence or absence of an ENF signal, determine the type of ENF, and verify the known date and time of a questioned recording by means of a one-to-one comparison of the evidentiary recording against a database recording. Automated analysis software allows for the automated comparison between the evidentiary recording and the database recording. This method is chiefly used to establish the date and time a recording was produced. In the case of both of these methods, it is important that all minimum requirements for the validation of ENF methodologies be met (Grigoras, Cooper, & Michalek, 2009:5).

Among the programs currently in use, Kantardjiev provides information on the two systems he recommends (2011:29-30):

- Sonic Advanced Wavelet Extractor (AWE) is a graphics processing unit (GPU) based sound analysis tool that is currently under development. The features of this software enable it to be suitable for locating and extracting disturbances that require further examination. The software has the ability to create a visual representation of the spectral contents of an audio file. It allows for the use of several different transform methods and to easily change between parameters such as window size, scaling function, etc. This provides a fast, smooth, and detailed method of analysis of the spectral contents of an audio signal. In experiments carried out by Kantardjiev, Sonic AWE has shown great potential to be used for both locating and extracting the ENF signal from a digital audio recording, as well as proving useful for other forensic applications.
- MATLAB Graphical User Interface (GUI) has been designed to simplify some of the processes used in audio forensics, such as filtering and searching for pattern matches within a database. The GUI provides a visual representation of the spectral contents of an audio file by means of a Short-Time Fourier Transform (STFT) in the time-frequency domain, as well as an FFT. The MATLAB GUI makes use of several linear-phase, low-pass, and band-pass filters to aid in the removal of unwanted frequencies – all with adjustable parameters. The software also allows for downsampling of the signal in order to reduce the computational requirements in processing the signal. The GUI is capable of extracting the ENF signal by various methods, including the STFT, the Capon filter-bank method (not addressed within the scope of this study), and measuring the zero-crossings of the waveform in the time domain.

When searching for matches in the database, it is possible to either load previously found patterns, or use newly extracted data. The user can define the amount of data that should be processed when searching the files. Searching for matches from the selected files is then done by specifying the frequency band from which the data was extracted, the maximum number of hits desired, and the threshold for what defines a hit (Kantardjiev, 2011:31). In his 2003 paper, Grigoras makes use of the program DC Live Forensics.

2.3.7 Residual ENF analysis

One of the criteria that must be met before presenting digital audio as evidence, is the validation of the integrity of the audio signal (Kajstura, Trawinska & Hebenstreit, 2004:165). It has been suggested that the analysis of the residual ENF signal could prove useful as a means of assessing the integrity of digital audio recordings (Rumsey, 2008:213). The ENF criterion is based on the analysis of the fundamental frequency of the electricity supply signal. If the recording device is powered by the main electrical network supply, and does not have an ideal voltage regulator, it is likely that the ENF signal will be recorded within the audio file as well as the intended audio. The detailed analysis of this electrical supply frequency allows for the investigation into the integrity of the audio recording and to possibly establish or verify the date and time when the recording was made (Kajstura, Trawinska & Hebenstreit, 2004:165).

Nominally this ENF signal would be steady at the supply frequency value of 50 Hz, but in reality the signal contains small random frequency variations over time due to the difference between electricity supply and demand. These unique variations over time provide the examiner with a unique timestamp that can prove useful in authentication examinations (Rumsey, 2009:745). The variations can be characterised through the detailed examination of a copy of the questioned audio recording that has been band-pass filtered around the designated frequency value (50 Hz) (Koenig & Lacey, 2009:684). Because of the random patterns inherent in the ENF signal over time, it is possible to determine the temporal authenticity of a recording, and investigate if the recording has been altered or manipulated. It may be possible to ascertain the time, date and location a recording was made by extracting the ENF signal and cross referencing the signal to that of a reliable database of stored ENF values (Rumsey, 2008:213). By comparing the recorded ENF pattern to a known database pattern from the same electric network, it is possible to verify or eliminate a questioned time of recording, and to determine the time of an unknown recording. In the former example it is possible for a technical specialist to conduct a visual comparison of the two extracted ENF patterns. However, in the case of the latter, it is necessary to use an automated software approach in order to efficiently search the database for matching ENF patterns (Huijbregste & Geradts, 2009:1). The technical specialist or examiner would then be required to verify the results of the automated analysis. In addition to these analyses, it may be possible to identify the type of power source used in the production of the recording (Koenig & Lacey, 2009:684).

It has also been suggested that battery-powered devices are able to record the ENF signal as a result of electromagnetic induction, provided the recording is produced in an electronically built-up environment or in close proximity to other mains-powered equipment, network transformers, or transmission cables (Grigoras, 2006:136).

The ENF criterion is based on two main assumptions: firstly, that the variation in the frequency value is the same throughout all points on the electrical network grid at the same point in time; and secondly, that the variations in the frequency are random and are not repeatable over long periods of time (Kajstura, Trawinska & Hebenstreit, 2004:166). In his 2006 paper, Grigoras determined that there was no predictable patterning in the variation over time around the 50 Hz frequency. The only possible exceptions to this are peak hours that experience more demand and maintenance operations or the switching of network components. Even as such, the values of the variation are not predictable, only their presence. Owing to the characteristics of electromagnetic wave propagation, the entire electric network which is formed by the interconnected systems including all the power stations (sources) and loads (shops, offices, houses, factories, etc) will have the same frequency, regardless of the point on the network from which it is measured (Grigoras, 2006:137).

This method of analysis may also help identify points at which an alteration was made to the audio file. In this instance, it may not be necessary to cross reference the file to an existing database (Kajstura, Trawinska & Hebenstreit, 2004:165). Since the ENF is a slow varying function of time, any abrupt changes in the frequency or phase of the frequency could be indicative of a discontinuity. This could imply an intended recorder operation (stop, start, pause), a manipulation or alteration to the file, or a damaged section of the file (Chang & Huang, 2010:178).

The methodology of ENF analysis requires the initial extraction of the ENF signal from the audio file by means of a band-pass filter, followed by the visual representation of the frequency as a function of time (Garg, Varna & Wu, 2012:67). A band-pass filter is applied to the audio recording centred around the designated supply frequency in order to isolate the desired bandwidth (Chang & Huang, 2010:178). The 3D-spectrogram (waterfall) and the 2D-spectrogram provide a continuous representation of the variation in the ENF signal over time (Grigoras, 2006:136). The same procedures for extracting the ENF signal are performed on a verified ENF database for the same network, date and time as the questioned recording. The examiner then compares the two signals over time, documenting any temporal and frequency differences between them. Any differences may indicate the presence of discontinuities or alterations on the submitted recording (Koenig & Lacey, 2009:684). The time that corresponds to the maximum similarity between the known and the questioned ENF signals is deemed to be the estimated time of production for the recording (Garg, Varna & Wu, 2012:67).

The applicability of ENF analysis is dependant on a number of factors, such as access to a reliable network database; identification of the true date and time of a recording independently of the embedded information; and the signal-to-noise ratio of the submitted recording (Koenig & Lacey, 2009:684). The ENF procedures may not be applicable to all recordings. The magnetic field coupling may be very weak or non-existent in some well-designed audio equipment, for example; when using condenser or piezoelectric microphones, or when using battery operated equipment remote from a noisy EMF environment (Maher, 2009:87). From a forensic view point, the continuing random variation of the frequency over time combined with its stability over distance (the network), that make ENF analysis a powerful resource for conducting authenticity examinations on recordings (Grigoras, 2006:137).

2.4. Power supply in South Africa

Power generation in South Africa is dominated by the state-owned utility, Eskom – who is responsible for approximately 95% of the electricity generated and supplied in South Africa. The department of Energy states that electricity is sold in bulk to the various municipalities, who then distribute it to consumers within their boundaries. The electricity supply is distributed in the form of alternating electrical current (AC) from the power supplier to the consumer via the national electricity distribution grid. This grid refers to the interconnected network of transmission lines and cables that connects the power stations throughout South Africa to the consumers. The national power grid refers to the combined transmission and distribution networks. Transmission lines are responsible for the bulk transfer of high-voltage electricity from the power generation facilities to the electrical substations, which are situated within the various municipalities. When transmission lines are connected to each other, they are termed a transmission network. The wiring that connects the local municipal substations to the consumers is known as the power distribution grid. Within their Quality of Supply Standards document (2003:14), Eskom differentiates between the allowed frequency deviations between a grid and an island network. Only the grid network values are referenced in this study.

Due to the thermo-electromechanical method of electricity generation and the equipment used to produce it, the supply frequency fluctuates randomly around the nominal supply frequency value (50 Hz). As consumers increase or decrease their electrical power requirements, there is an imbalance between the electrical system load and the system generation. These mismatches result in transient changes in the supply frequency of the electrical network (Gerazov, Kokolanski, Arsov & Dimcev, 2012:1164). In order for the frequency to remain within the allowed deviations, the power plant has control measures in place. Since these control systems are mechanical operations, there is an inevitable time delay for the corrective measures to be evident at the end user. This results in the unique time-varying deviations around the supply frequency (Chang & Huang, 2010:176). Since power cables directly connect the grids, this implies that the variations around 50 Hz are evident at the end user, and that they are uniform throughout the grid. Because of the complex behaviour of the many consumers connected to the power grid, the variations in the ENF can be considered unique and unpredictable, and thereby suitable to be used as a unique timestamp for a given time period (Gerazov, Kokolanski, Arsov & Dimcev, 2012:1164).

The Eskom fact sheet (2013) states that the national control centre at Simmerpan in Germiston controls the power transmission network throughout the country and is aware of what the average demand is for South Africa.

2.4.1 Electricity generation and supply

Power plants produce electricity to be distributed according to the designated frequency supply as defined by the National Energy Regulator for the specific country (in South Africa, NERSA). It is a requirement that all devices powered by alternating current (AC) and connected to the main electrical network are compliant with this standard (Chang & Huang, 2010:176). The Eskom Fact Sheet (2013) confirms that in South Africa, electricity is generated and transmitted as alternating current (AC) at a nominal supply frequency of 50 Hz. The generators are synchronised to the National Grid at a frequency of 50 Hz.

The rotational speed of the generator rotor, and the variations in the heat source and pressure are responsible for the random frequency deviations in time (Chang & Huang, 2010:176). In their fact sheet (2013), Eskom confirms that there is a strong correlation between the rotational speed of the generator and the frequency of the electrical supply. Most generators within the Eskom system are classified as two-pole generators. The rotor of a generator is a large electro-magnet with a pair of opposing poles. The rotational speed of these generators is nominally 3000 rpm. When the rotational speed of a two-pole generator is 3000 rpm, the frequency over time (3000 divided by 60 seconds) is a resultant 50 Hz.

The frequency of the supply can be a good indicator of what the demand for electricity is at any given moment. A decrease in the load (less demand) causes the speed of the generators to increase, causing a rise in the supply frequency. It is the responsibility of the control systems at power stations to ensure that the supply of electricity from the generators is reduced in order to lower the frequency back to 50 Hz. Likewise, an increased load (increased demand) on the network causes the rotational speed of the generators in the power plant to decrease, thus resulting in a dip of the supply frequency. In this case, the output from the generators must be increased to meet the new demand (Eskom fact sheet, 2013). The power supplier must compensate for the effects of the load changes (Kantardjiev, 2011:7).

2.4.2 Quality of supply standards and frequency deviations

According to the NRS 048-2:2003 Quality of Supply Standards document (Standards South Africa, 2003:13), the standard frequency of supply in South Africa is 50 Hz. The nominal deviation from the standard frequency allowed within a grid network is ±2% (± 1 Hz). The maximum limit for the variation of the frequency of the supply voltage in a grid network is ±2.5% (±1.25 Hz).

Should the frequency drop below a certain threshold limit, automatic load shedding will occur. In order to prevent this, Eskom performs controlled load shedding so that the network and all electronic devices and appliances connected to the network are protected against under-frequency. Controlled load shedding is a way of distributing the available electricity between all customers when the demand is greater than the supply. By controlling the amount of demand by means of load shedding, the balance between the supply and demand is restored, and the frequency of the supply is maintained at a suitable level (Eskom fact sheet, 2013).

The national power provider records all interruptions in the power supply for the period of one year. The most common interruption events are:

- Forced operational interruptions;
- Voluntary customer load reductions;
- Planned interruptions; and
- Involuntary customer load interruptions.

Interruptions can be further classified as either momentary interruptions or sustained interruptions (Standards South Africa, 2003:23).

2.4.3 Database of Supply

Since the ENF is the frequency at which the voltage levels in an electric network oscillate, it is possible to obtain the ENF pattern by analysing the voltage level signal directly from a power socket outlet (Huijbregste & Geradts, 2009:2). In an ideal situation, signal from the electrical network would be free of disturbances. However, this is not the case in reality. There are several factors that affect the accuracy of the measurements (including harmonics), these are voltage spikes, transient spikes, and other components from within the network that may cause disturbances in the supply frequency. Since only the 50 Hz frequency is of interest in this application, a band-pass filter is applied in order to isolate the required frequency and minimise possible errors in the measurements (Kantardjiev, 2011:26).

Storage of ENF data can be done by sampling the signal at intervals and smoothing out some of the noise. Due to the fact that the signal frequency is low compared to typical audio signals, the signal can be downsampled. This will minimize the processing requirements when extracting the signal. However, in order to retain detailed information about the harmonic structure of the signal, it may be necessary to create a spectrogram of the signal prior to downsampling (Rumsey, 2009:745). Brixen suggests that it may be possible to obtain the ENF logging data from the electricity supplier, but that the data is usually only kept for a few months. Thereafter, only the statistics of the supply are kept, as well as any information regarding abnormalities in the supply (such as dropouts in a particular area) (Rumsey, 2008:216). Possible sources of historical ENF data could be (SWGDE, 2010:10):

- Power generators;
- Power transmitters;
- Independent Service Providers;
- The National Electricity Regulator;
- Universities; and
- Other national laboratories (e.g. CSIR).

2.4.4 Properties of a reference database

In order for a reference database to be considered credible, it must maintain its integrity. As such, it may be necessary to develop validation protocols to ensure the integrity of the data. It is important that the reference data set make use of a reliable external time synchronization method (either by means of web-based time-servers or GPS satellites). Any possible date or time synchronisation errors must be measured and taken into account when conducting analyses using the reference data set (SWGDE, 2010:5). In order to minimise the effect of bias, a high quality external clock reference should be used in conjunction with a laboratory instrument grade computer audio interface (FSAAWG, 2009:5).

2.4.5 Methods of setting up and maintaining an ENF database

When recording or collecting ENF data as a reference source, the following points must be taken into consideration (FSAAWG, 2009:8):

- The ENF data collected must not be compressed.
- Regular backups should be made of the ENF database.
- Two separate databases should be maintained in different locations in order to reduce the risk of accidental data loss in the case of network power failure or measuring equipment failure.
- Do a regular verification of the system date and time and manually make any necessary changes where no external synchronization source has been implemented.
- Isolate and employ the necessary security protocols to protect the system from unauthorized access.

The Scientific Working Group for Digital Evidence (2010:4-5) recommend the point of collection for the ENF reference database be either at the transmission point or the distribution point. The transmission point refers to the source of the electricity supply, whereas the distribution point refers to the supply present at the end user (wall socket). The benefit of collecting your data from the transmission point is that the frequency is unaffected by interference at the local grid level. However, it is very difficult to access these points. The advantage of collecting data at the distribution point is that the frequency and voltage are ubiquitous throughout the grid. The frequency can, however, be affected by local sources of interference (e.g. circuit breaker trips, and load shedding). It is important to remember that the voltage is different at these two points, and it is imperative to use suitable equipment in order to avoid incorrect data or damage to equipment and data. It is suggested that data be collected at multiple distribution points throughout the grid in order to reduce the effects of local interference. A robust reference database collection system will make use of multiple sources in order to prevent against signal loss as a result of interference (such as circuit breaker trips or line failure).

The following are some points of concern when making use of archived ENF data (SWDGE, 2010:11):

- What is the format of the data?

- How long is the data stored?
- Is it publicly available? If so, what are the security implications?
- What are the conditions for use, if any?
- Are there any liability issues or legal concerns?
- Are there any other issues or concerns that should be addressed?

Access to, and maintenance of, a credible and reliable database is essential for the success of ENF analysis as an audio authentication tool.

2.5 Procedures and methodologies of ENF analysis as an authentication tool

In the context of audio forensics, the application of ENF analysis seeks to answer some of the following questions (SWGDE, 2010:4):

- Was the recorder connected to the main electricity supply?
- Is the recording a duplicate?
- Is the recording a continuous representation of the acoustic event?
- If multiple sections are present, were they recorded within the same electrical network grid? Within what time frame were the multiple sections recorded?
- On what date and time was the recording produced?
- Where was the recording produced?

2.5.1 File preparation

In order to reduce the risk of errors in the analyses, it is necessary to prepare the working file adequately for optimal playback. Optimal playback ensures that the critical listening and subsequent examinations are conducted with minimal playback artefacts. For many of the examinations, conversion of the file to a standard, uncompressed audio file format such as a PCM wave file is required for the production of visual representations such as waveforms and spectrograms (Koenig & Lacey, 2009:671).

Optimal playback begins with determining the format of the data on the medium and identifying the proper system and software for playback. For the purposes of digital audio analysis, media are usually submitted in four types of categories (Koenig & Lacey, 2009:672):

- *Standard removable media*. This includes CDs, DAT, MiniDisc, MiniDV and high definition DVD formats (Koenig, Lacey & Killion, 2007:362).
- *Standard file formats*. This includes Pulse Code Modulation (PCM) wavefiles (.wav) and compressed audio formats such as MP3 that do not require the use of proprietary software for playback. These formats can typically be reviewed with the use of conventional audio editing or playback software. These may be presented on any number of media (Koenig, Lacey & Killion, 2007:362).
- *Proprietary file formats*. This includes any specialised formats which require the use of unique software or hardware for playback, conversion or access (Koenig, Lacey & Killion, 2007:362).
- Fixed memory devices with no digital interface capability.

The optimal method of playback depends on the format of the file and the properties of the recorder. There are three options for playback of standard and proprietary format files (Koenig, Lacey & Killion, 2007:362):

- Conversion of the file to Red Book-compliant standard and playback on a stand-alone unit.
- Direct playback of the file in applicable software via the audio interface of a computer.
- Playback directly from the digital device.

The Red Book-compliant option often yields the best results since it is unaffected by the audio interface characteristics, computer irregularities, and software limitations.

A bit-for-bit copy or digital image of data contained on the medium is made and archived appropriately. Specialised testing confirms the accuracy of the copy compared to the original. Any analyses of the recording should then be conducted using the copy of the file. The original should only be accessed if and when another copy or digital image is needed (Koenig, Lacey & Killion, 2007:362).

2.5.2 Extraction of the ENF signal

Changes in the ENF may be very small, and may require the use of relatively high-resolution spectral analysis equipment (FSWAAG, 2009:8). Since the frequency deviations to be examined are of such small magnitude, they are easily corrupted even at fairly high signal-to-noise ratio (SNR) (Kantardjiev, 2011:43). In addition to noise, problems with DC offset may affect the accuracy of the estimations (Kantardjiev, 2011:22). Since all electronic devices and systems contain noise sources of some kind, it is important to only use high quality sampling devices with very low inherent noise levels and distortion to obtain accurate analysis results. Grigoras (2003:3) recommends the use of a good quality audio interface with a signal-to-noise ratio less than -94 dB, and a total harmonic distortion below 0.003%.

The conditions of the recording as well as the requirements of the analysis will determine the most suitable method of data extraction and analysis. For example, under certain conditions, the time domain analysis may be better suited and provide more reliable results than the frequency domain method. It is important to select the correct analysis method and follow the practices for that technique.

Liu, Yuan, Markham, Conners and Liu (2011:4), and Huijbregste & Geradts (2009:3) recommend downsampling the signal (from 44100 Hz to approximately 300 Hz for a 50 Hz frequency) before any of the other steps are carried out. This will reduce the computational time for signal processing. Also suggested is the use of a low-pass filter before downsampling the signal in order to reduce the effects of digital aliasing. The following are a few of the recommended prerequisite processes for the recording prior to extraction and analysis to ensure reliable and repeatable results (FSAAWG, 2009:7):

a) Removal of DC components.
b) Application of bandpass filtering in order to remove signal outside of the designated ENF bandwidth. (This should also remove any DC components present in the audio signal).
c) Downsample at a frequency of twice the ENF signal plus 40% (ENF x 2.4). Therefore, for 50 Hz, it is necessary to downsample to 120 Hz.

The method of analysis will determine the steps that follow the initial processing.

There are currently three main methods for the analysis of the ENF signal (Liu, Yuan, Markham, Conners & Liu, 2011:4):

- Time/frequency domain analysis - based on the spectrogram
- Frequency domain analysis - based on the calculation of the maximum magnitude of a series of power spectra
- Time domain analysis - based on the zero-crossing measurements

In his 2009 (2009:646) paper, Grigoras defines the three main methods of analysis as follows:

- The **time/frequency domain method (spectrogram)** involves computing spectrograms of the ENF signals for both the questioned recording and the database recording, and conducting a visual examination and comparison of the two. This method is the fastest and easiest to implement. It can clearly reveal the total number of ENF signals present in an audio file, and is especially useful when cross-referencing questioned to known ENF signals for temporal verification. It is also practical for recordings that exceed the duration of 10-15 minutes. Grigoras recommends that the spectrogram method be used first to determine the number of ENF components present, and provide information that may help determine the following analyses.

- The **frequency domain method** involves computing FFTs over short time windows, extracting the maximum magnitude value around 50 Hz, and comparing the questioned samples to that of a known ENF database set. The frequency domain method is usually applied on recordings that contain more than one ENF component to separate the frequency signals in order to identify the date and time each different component was produced. This method is used when cross-referencing the ENF of a questioned recording against an ENF database to check the integrity of the recorded audio, and also to identify the date and time the recording was produced.

- The **time domain method** consists of calculating the zero-crosses measurement and comparing the questioned ENF component to that of the ENF database. This method is limited to recordings that contain only one ENF component. This is proposed to be the most in-depth analysis method, offering more information than the frequency domain method under certain circumstances. When a questioned recording is compared to a database recording in the time domain, non-identical ENF shapes can be obtained as a result of sample synchronisation delays. To reduce the possibility of these errors, it is recommended that any DC components be removed from the recording and a 50 Hz band-pass filter be applied. By applying a band-pass filter between 49-51 Hz, without downsampling the frequency, it is possible to separate the ENF signal from the rest of the recording. Particular attention must be paid to averaging operations over windows of different lengths. This method displays the micro-variations of the ENF signal by plotting all the zero-crosses of the residual ENF signal, and measuring the time differences between the consecutive zero values.

In order to evaluate non-stationary signals, the signal must be divided into time frames and processed separately to be able to view the variation of the frequency contents over time. It is, however, difficult to obtain high frequency resolution and high time resolution at the same time. Processing samples of long time duration can yield a high frequency resolution where closely adjacent frequency components can be separated from each other, but with less information available regarding the frequency deviations in time. The inverse also applies: processing shorter time domain samples provides a lower separation of close frequencies, but more information about the frequency variations in time (Kantardjiev, 2011:20). Grigoras (2003:1) concluded that the waveform and the FFT spectra are useful to view the periodicity of the ENF over short time windows. But the most applicable methods of examining the information and variation of the ENF frequency over time are the waterfall (3D-Spectrum) and 2D-spectrogram methods. These two methods reveal the history of all peak values in time, allowing for a clear view of the variation of the ENF signal.

2.5.3 Database comparison

It is necessary to conduct an initial analysis on the questioned recording to ascertain the presence and condition of an ENF signal. If this analysis suggests that the recording contains a suitable ENF signal, and the resolution is suitable for comparison, the examiner may begin extracting the signal for cross-referencing against a known database (Rappaport, 2012:43). Both sets of ENF data (questioned recording and database) must be directly comparable with matching the time and frequency resolutions. The compatibility of the two sets of data being used for comparison is essential for reliable results. For example, if the database stores the file in a 16-bit, single channel .wav audio file with a sampling rate of 120 Hz, it is important that the questioned signal be converted to the same format. This is essential for automated software analysis (FSAAWG, 2009:8). When analysing the differences between the questioned and reference database ENF values, it is important to make sure that the FFT values used in the extraction of both data sets are the same. The number of points analysed for the FFT influences the accuracy of calculating momentary frequencies.

This will cause a difference between the real value of the momentary frequency and the value extracted from the recording. However, tests did conclude that this limitation was overcome if the duration of the questioned recording was sufficiently long (10-20 minutes) (Kajstura, Trawinska & Hebenstreit, 2004:168).

In their ENF discussion paper, the scientific working group for digital evidence (SWGDE, 2010) recommend the development and implementation of an evaluation protocol in order to produce meaningful analysis results. Robust testing will define the parameters, as well as ascertain the strengths and limitations for each technique. Testing will entail systematically evaluating the effect of all variables on the results, such as sample length, window length, window overlap distance, signal to noise ratio, relative harmonic strength, original sampling rate, etc.

When comparing two ENF data sets for temporal authentication, it is important to take the natural cyclical patterns within the network into consideration in order to avoid false conclusions. The following are examples of cyclical patterns (SWGDE, 2010:6):

- Load following (minutes)
- Scheduling (hourly)
- Load behaviour, time correction (daily)
- Weekday vs. weekend (weekly)
- Special days (holidays, sporting events)
- Seasonal trends (lighting, climate control)

In order to locate the time of a recording, an extracted ENF pattern must be matched to a pattern in the reference database (Kantardjiev, 2011:24). The two main methods of comparison are visual (spectrographic) and semi-automated (FFT). The former relies on finding visual consistencies between spectrograms of the ENF signal from both the questioned recording and database. This method can be very difficult and time consuming unless there is a suspected discontinuity or anomaly at a known location within the recording that requires further analysis. Where the signal level is very low and cannot be improved by means of signal processing, this may be a useful method of analysis. However, a recording that suffers from a low signal to noise ratio may not benefit from this technique (Rappaport, 2012:46). Kantardjiev (2011:24) suggests the use of an additional method, i.e. comparing the distance measurements between the ENF patterns of the recorded audio and the database. By measuring the distance between adjacent points in two data sets, the sum of the distances can be used as a measurement of how similar they are. The minimum distance should represent the best fit between an extracted pattern and the reference database. When measuring the distances between the zero-crosses, it is important that the examiner be aware of the effects of DC offset on the ENF signal. The offset changes the spacing of the semi-periods and can lead to false conclusions (Rappaport, 2012:46).

2.5.4 Interpretation and evaluation of the results

The interpretation and evaluation of the analysis results must take the following into consideration (FSAAWG, 2009:9):

a) Available information regarding the background and circumstances of the recording
b) Original expectations and requirements formulated during the case assessment
c) Other possible audio analysis techniques

It is important that the examiner have a good knowledge of the circumstances in which the recording was produced; the recording could have been made in a building that uses its own generators for power supply. Depending on the source of power, the fluctuations in the ENF may differ in nature. These differences could provide some information about the potential source of the audio data (Rumsey, 2008:215-216).

There is a strong likelihood that a recording that has more than one ENF component is not authentic, or that it may have been edited. Separation of the components by means of the spectrogram method is suggested to determine the actual number of ENF components. The individual frequency components can then be extracted and analysed to determine the time at which the segments were added.

If there is a good degree of consistency between the questioned recording and the database recording, the likelihood of verification that the file is continuous and unaltered is strong (Rappaport, 2012:44). If the two ENF signals have definite segments that match, but some segments that do not match and/or disrupt the time sequence, it is a strong indication of manipulation. This is especially so if the time sequences are out of chronological order (i.e. sections have been "cut and pasted"), repeated, or originally from different times and dates entirely. Identifying the different times and dates would require multiple ENF comparisons (Rappaport, 2012:45).

Segments of the ENF signal extracted from the questioned recording that differ in time duration from the database, but do not change the basic chronology or introduce an ENF signal from other dates, times or sources, may be a result of stops or starts made by the recording operator at the time of production. The questioned recording ENF may appear to have skips and be a series of segments linked together. A file where the original ENF signal has been replaced by an alternative/unaltered ENF would likely show traces of an external editing DAW or re-recording. It could also contain traces of the original ENF harmonics if they were recorded (Rappaport, 2012:45).

According to Rappaport (2012:44-45), an anomaly or discontinuity within an audio file is not always indicative of an alteration, e.g. where there is a high degree of correlation between the questioned recording and the database recording, but there are momentary spikes in amplitude present in one file, but not in the other. These spikes are most likely a result of local voltage fluctuations or power surges and not tampering – provided that the 50 Hz frequency itself is not altered. Another possible explanation for missing samples or short variations in one file and not the other could be disc write errors. If the ENF signals of the two files being compared begin at the same point in time, but slowly diverge over time, it could be the result of an unstable digital word clock and not tampering. If the ENF signal is present during some sections on the questioned recording, but not others, it may be due to a change in the power source e.g. laptops that can change from mains to battery power (there would normally be a change in the DC offset in this instance). In his 2009 study, Michalek found that some recorders showed a shift of the ENF on the frequency axis, independent of the compression levels applied to the signal. This is most likely due to the quality of the components used. This signal still allows for ENF authenticity analysis.

According to Grigoras (2009:656-658), the main sources of errors encountered are:

- Wrong system digital clock timing settings. This can lead to erroneous acceptance or rejection of the evidence or to false identification of the date/time. It is important that the database clocks be checked regularly and synchronised with an atomic clock. Many low quality digital recorders will also contain faulty digital clocks. It is important that the digital clock of the recorder in question be analysed if possible.

- Any DC offset must be removed from the questioned recording, as it can lead to incorrect zero-crosses measurements.
- Noise present in the recording can also affect the zero-crosses measurements. Noise can be removed by applying a band-pass filter with the cut-off frequencies below 49 Hz and above 51 Hz. This filter removes any other non-relevant signals but leaves the ENF signal intact and unaffected for database storage or analysis.
- The FFT order level can affect the accuracy of analysis results from the frequency domain method.
- It is important to ensure that the examiner is working with the same averaging operations over different window lengths for questioned vs database cross-referencing.
- Distortion or clipping of the original questioned recording signal can inhibit the process of extracting and analysing the ENF component.
- The clocks, quantization and sample rates all affect the quality of the ENF signal. The recording equipment should be analysed if possible to determine the effects of equipment induced artefacts on the signal.
- Any secondary ENF signal induced in the copying process may mask the original ENF signal.

2.5.5 Effect of an electromagnetic field on battery-powered devices

When an electric current flows through conducting materials, a magnetic field is produced that is directly proportional to the magnitude of the current. As such, it is assumed that when the main network powers equipment, it will create an electromagnetic field with a frequency of 50 Hz. There is also evidence of the magnetic field existing at the same frequencies as the power harmonics (Kantardjiev, 2011:11). When battery-powered electronic equipment or microphones are placed in close proximity to an AC-powered electromagnetic field (EMF), there may be an audible hum induced into the recording at 50 Hz (Grigoras, 2009:653). Surrounding environmental factors such as fluorescent lighting could induce this. This hum could be indicative of the ENF signal or one or more of its harmonics (Koenig & Lacey, 2009:676).

Experiments conducted by Kantardjiev (2011:43), Michalek (2009:357), and Grigoras (2009:653) show that it is possible for disturbances related to the ENF to be captured in recordings produced on battery-powered devices, provided that the recording was produced in an area well covered by the electrical network. Further tests conducted by Kantardjiev, Michalek and Grigoras concluded that an increase in the distance between the recording system and the EMF source resulted in a decrease in the magnitude of the EMF signal induced into the recording. Grigoras approximated that at a distance of about 2 m, the real ENF induced into the recording becomes too difficult to estimate. As a result of the application of systemic audio compression, these devices are more susceptible to recording the higher frequency harmonics of the ENF rather than the fundamental 50 Hz (Kantardjiev, 2011:44).

The susceptibility of the recorder to capture the ENF signal is not proportionate to the magnitude of the magnetic field. Some less expensive recorders did record the ENF signal, but at a slightly shifted centre frequency (Rumsey, 2009:745). Some Faraday shielded battery-powered devices were more resilient to the interference of the 50 Hz signal (Kantardjiev, 2011:44).

Results of the experiments mentioned above showed that none of the test recordings indicated a strong disturbance around 50 Hz, but there was an indication of an ENF signal at 100 Hz (likely the harmonics of the ENF signal). All recordings showed uniform disruptions that coincided with the equipment producing the EMF being switched on and off. Extraction of ENF components from these recordings proved more challenging than from mains-powered devices. Although none of the methods proposed worked well for this form of analysis, the STFT was shown to be the preferred method of extraction, resulting in the least disrupted pattern (Kantardjiev, 2011:44).

2.5.6 Conclusions from previous studies

Studies carried out at the Institute of Forensic Research confirmed that the ENF criterion could be used to investigate the authenticity and estimate the time of the production of a recording (Kajstura, Trawinska & Hebenstreit, 2004:171). The results of further studies carried out by Kajstura, Trawinska & Hebenstreit (2004:168) confirmed the various hypotheses outlined in previous sections. The momentary value of the frequency is the same throughout the network grid: the frequency variations over a long period of time are random in nature and can thus be the basis for establishing temporal authenticity of an audio recording. It was also determined that the ENF criterion could be used to identify points within an audio file that have been altered, as well as possibly even establish the nature of the manipulation. In the case of a cut/deletion, it was possible to estimate the time duration of the fragment that was removed. If the recordings used to manipulate the file were made under different conditions or using multiple recording devices, it is likely that the combined fragments will differ significantly in terms of the ENF component (Michalek, 2009:359).

All the methods described have been proven suitable for extracting the ENF patterns where there is no strong interference by noise. Frequency domain analysis and the zero-crossings methods produced smoother pattern estimates at low frequencies in comparison to the STFT (Kantardjiev, 2011:43). However, frequency domain analysis requires extensive computational processing for the frequency extraction. The STFT has also been proven to be a fast and reliable method for estimating the spectral contents of an audio file. The zero-crossings method was found to be the simplest method, providing the most frequency information in relation to the computational effort required. This method can provide useful information regarding the near instantaneous frequency deviations and reveal shorter alterations in a recording than is possible by the other methods (at the suggested settings). However, the zero-crossings method is not recommended for recordings that contain noise. The induced noise produces incorrect estimates near the crossings, resulting in inaccurate analysis results (Kantardjiev, 2011:43). Apart from the zero-crossings method, the other methods listed are able to withstand a large degree of noise present in the reference signal before the results are affected. However, in order to provide reliable analysis results, the extracted pattern must be fairly intact, and the signal may not fall below a certain SNR. Although the general pattern of the ENF may still be present in a signal with a SNR below 10 dB, it is not recommended to use the signal for analysis as the methods proposed in this mini-dissertation tend towards inaccurate results (Kantardjiev, 2011:38).

It is possible to detect locations where an audio file may have been manipulated by means of the visible spectral leak effect. This presents as a visible blur of the frequency concentrated near the fundamental frequency (50 Hz) that overlaps into nearby frequencies, disappearing into a very narrow band where the discontinuity occurs (Michalek, 2009:359).

2.6 ENF analysis as an authentication tool

ENF analysis and its application as an authentication tool have been studied and promoted by many of the prominent professional and forensic groups, namely the (Liu, Yuan, Markham, Conners & Liu, 2011:1):

- Audio Engineering Society (AES)
- Scientific Working Group on Digital Evidence (SWGDE)
- American Academy of Forensic Sciences (AAFS)
- European Network of Forensic Science Institutes (ENFSI)
- Forensic Speech and Audio Analysis Working Group (FSAAWG)

The ENF criterion aims to authenticate digital audio according to the main assumptions of an authentic recording as set out by the Audio Engineering Society; relating to the integrity of the recording and its having been recorded simultaneously with the course of events in question (Michalek, 2009:355).

ENF is suitable for forensics due to the fact that (Chang & Huang, 2010:176):

- The ENF is distributed only within the interconnected grid – allowing for localization of the recording.
- The ENF is time varying – allowing for temporal identification.
- The ENF is generated by the power plant – a reliable source for the ENF.
- In the absence of an ENF database, it is possible to analyse the ENF in order to detect discontinuities.

By making use of the ENF signal as a method of authentication, the signature used for the authentication examination is independent of the recording device, as opposed to the signature for analysis being a part of the individual device (such as the electromagnetic components of the microphone) (Chang & Huang, 2010:176).

2.6.1 Limitations of ENF analysis as an authentication tool

Each method of extraction and interpretation has its own limitations and conditions of use that should be well understood, as well as compatible with the reference database. It is important to validate and compare the various methods of ENF analysis before their value as a legal method of authentication can be determined (SWGDE, 2010:6).

The opportunity for detection and recovery of ENF data depends on multiple factors, including (FSAAWG, 2009:6):

- Type of power supply used by the recorder during the production of the recording
- Technical specifications of the recorder including:
    - file format (WAV etc.)
    - compression algorithm (MP3, WMA, etc.)
- Signal-to-noise ratio of the recording
- Location of the recorder relevant to the source
- Age and condition of the media
- The level of any security features applied by the user

The length of the registered ENF signal on the recording has a direct influence on the accuracy of the temporal authentication of the recording. The level of interference (SNR) and the length of the window function used to calculate the FFT functions also impact the accuracy of the analysis. For recordings that have a low SNR, a longer length of recorded ENF signal is required for accurate authentication (Michalek, 2009:358).

Grigoras tested the effects that different compression algorithms had on the ENF signal. The experiments revealed that most compressed ENF recordings could be used for spectrographic comparisons against a database. However, it was also deduced that the compression algorithms could mask editing marks and affect the signal-to-noise ratio, thereby affecting the compressed ENF shape (Grigoras, 2009:654). Some low bit-rate compression schemes such as 8 Kb/s MP3 managed to preserve the basic ENF component information when recording audio data (Rumsey, 2008:217). It is therefore recommended that the recording equipment be characterised prior to audio analysis in order to determine the effect of its compression algorithm on the ENF signal (Grigoras, 2009:656).

The reliability of ENF analysis is limited by similarities within the techniques used to extract the ENF pattern from the database that is used for comparison. ENF analysis is also hampered by the presence of a frequency offset, resulting in erroneous matches (Huijbregste & Geradts, 2009:3).

CHAPTER 3

METHODOLOGY

This validation process addresses the main hypothesis of this mini-dissertation, namely: To what degree can ENF analysis be used as an authentication tool in South Africa. In order to verify the questions posed in this mini-dissertation, the following methodologies were employed:

- An extensive literature review investigates the principles and fundamental aspects of ENF analysis.
- An empirical validation of the applicability of ENF analysis within a South African context.

The empirical validation addresses the main research question in terms of:

- The possibility that the ENF can be recorded into, and extracted from a digital audio recording;
- The effect of microphone characteristics on recording the ENF signal;
- Infrastructure required for ENF analysis;
- The feasibility of establishing and maintaining an ENF database;
- The extraction of the ENF signal from a digital audio recording and comparison of the extracted signal to a known database of ENF values;
- The effectiveness of ENF analysis for detection of discontinuities/disturbances; and
- The required conditions for ENF analysis, including the limitations of the results.

To ascertain the applicability of ENF analysis, the empirical validation addresses the various sub-questions posed in Chapter 1.5. The procedures followed for testing the hypothesis are outlined below.

3.1 Recording and extracting the ENF signal

In order to determine whether an ENF signal can be detected, a 16 bit, 44.1 KHz digital audio recording is made with, and without audio input. For this purpose, the equipment used to record the audio is powered by the electric network, without utilising a UPS device or voltage regulator that could affect the integrity of the ENF signal.

The recording is digitally processed to establish the existence of the ENF signal at 50 Hz. The processing of the audio file entails the application of a bandpass filter centred at 50 Hz, a cut off filter with a 150 dB/octave slope in order to exclude additional signal that could interfere with the ENF signal, followed by a high-resolution spectrogram. This is produced by means of a Fast Fourier Transformation or FFT (FFT buffer size of 4096, Butterworth Filter, Hanning window, and a window overlap of 50%). The sampling rate of the signal is reduced from 44.1 KHz to 150 Hz after applying the bandpass filter to reduce the processing requirements on the computer.

The initial processing is done on a recording that has not been manipulated, and that contains mostly silence and some speech. This serves the purpose of determining the possibility of recording, extracting, and analysing the ENF signal. The recording analysed is longer than 30 minutes in order to allow adequate time for the deviations in the supply frequency to be evident. The purpose of analysing a recording with some speech, and some silence is to determine the obscuring effect of acoustic information on the detection of the ENF signal. All the files are recorded without using any audio filters, audio amplitude attenuation pads, or audio dynamics compression algorithms.

The aims for this experiment are:

- To determine if the ENF signal can be detected in an audio recording that has been recorded by a device powered by the electrical network;
- To determine if the ENF signal (if present) can be extracted for further analysis; and
- To determine the effect of additional acoustic input on the detection and analysis of the ENF signal.

The extracted ENF signal is also utilised in order to address some of the subsequent questions posed within this study.

3.1.1 Influence of microphone characteristics on the ENF signal

In order to determine the influence of the microphone characteristics on the accuracy of the recorded ENF signal, an audio recording is made with types of microphone as the variable. The variety of microphones chosen includes dynamic, large diaphragm condenser and pencil condenser microphones. The same external audio interface is used to simultaneously pre-amplify all of the different microphones on individual channels for recording on the DAW (powered by the same electrical outlet). This eradicates the question of variability as a result of time differences or the properties of the external audio interface.

Due to availability of equipment, this aspect of the study is conducted at the recording studio at St Mary's Diocesan School for Girls, Hatfield, Pretoria. The following equipment is used:

- MacBook Pro
- Focusrite Saffire Liquid 56 external audio interface (firewire)
- Rode NT 2-A (large diaphragm condenser microphone)
- Rode NT 1-A (large diaphragm condenser microphone)
- Samson C02 (pencil condenser microphone)
- Shure Beta SM57 (dynamic microphone)

The microphones connect into the audio interface from channels 3-6 in the respective order as listed above. The Class A pre-amplifiers of channels 3-6 of the Saffire interface are utilised. The software used to record the audio files is Logic Pro version 10. For the purposes of this study, all audio is recorded at a sampling rate of 44.1 KHz, and a 16-bit bit depth.

In order to gain perspective of the deviations of the ENF values over time, the duration of the recording is one hour. All unnecessary equipment such as the air conditioner and audio monitors were turned off to avoid any interference with the recording.

3.2 Setup and maintenance of an ENF database

There are two proposed possibilities for recording a database of the ENF values:

1. Record an audio file without acoustic input, and then extract the ENF signal. The extracted ENF signal data is stored as the ENF database.
2. Utilising a data acquisition unit that records the voltage, to log the frequency data directly from a household electricity distribution point (electrical socket). The data would ultimately be presented as frequency as a function of time.

Both of these possibilities are investigated to ascertain their respective strengths and weaknesses as techniques for the creation of an ENF database. Both options capture a continuous measurement of the ENF signal from the distribution point. The benefit of recording from the distribution point is that the frequency values that are distributed to the end user are recorded, complete with deviations, and any interference that may have occurred within the local network itself. The purpose of the logged data is for comparison against digital audio recordings for authentication; as such, it is necessary that any interruptions that may occur within the electrical supply be documented to ensure the accuracy of any later analysis.

3.2.1 Recording the frequency as an audio file

In order to assess this option, a digital audio recording will record silence continuously for approximately 24 hours. The proposed equipment for this experiment is a microphone, external audio interface, computer and DAW. The recording is subsequently processed to isolate the ENF signal. Conclusions regarding the suitability of this technique for recording a database signal are reported.

Following is the suggested process for extracting the ENF signal from the audio file. The details proposed in this section are as specified by the Best Practise Guidelines for ENF analysis (FSAAWG, 2009:7-8), and are taken as a starting point for extraction:

- Bandpass the signal
- Downsample
- Compute an FFT
- Create a frequency vs time plot

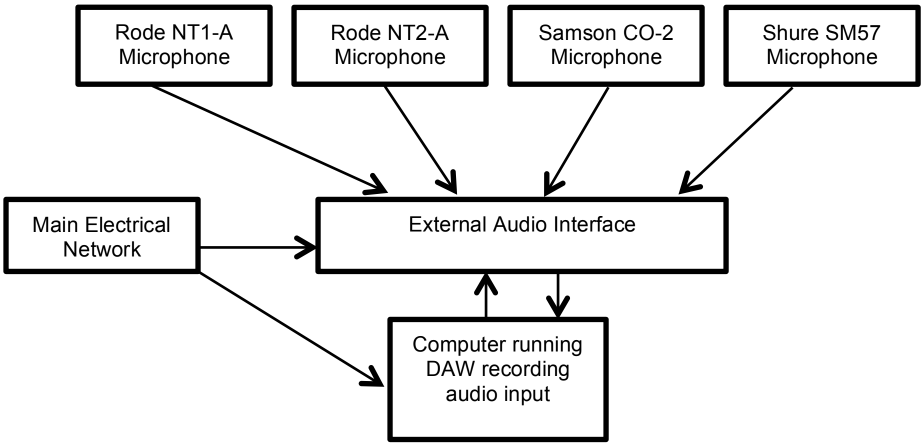

Diagram 1 – Experimental setup of equipment utilised to record the ENF database as an audio file.

3.2.2 Recording the frequency via a data acquisition unit

Proposed equipment for the logging of the ENF data is a data acquisition unit, with a step down transformer, and a computer with the necessary software to process and store the data. It is possible to apply a bandpass filter in order to isolate the 50 Hz frequency and thereby reduce any interference of other unwanted signals. However, it is important that no aliasing effects or any other disturbances on the ENF are induced into the signal. As such, an anti-aliasing filter that is incorporated into the data acquisition unit is applied with a cut off frequency of 60 Hz. The data is measured at a sampling rate of 150 Hz. This ensures that the limits of the frequency are still captured, but excludes any unnecessary frequency data that could interfere with the recorded ENF signal.

Equipment used for recording the database via a data acquisition unit:

- Voltage divider circuitry built in to a 3 point power socket
- BNC coaxial cable
- Data acquisition unit (measures samples of the voltage)

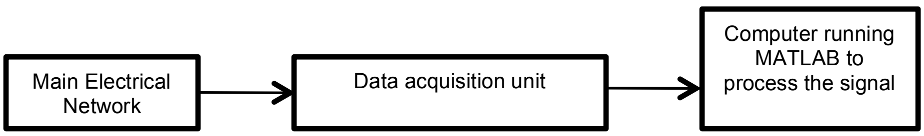

Diagram 2 - Experimental setup for recording the ENF database utilising a data acquisition unit.

3.3 Comparison of extracted ENF signals

In order to determine whether an extracted ENF signal can be compared to the recording of a known ENF database (and establish the accuracy of the results), a test audio recording is produced at the same time the database values are recorded by the data acquisition unit. The recording will have the same format characteristics as the database, i.e. 44.1 KHz, 16-bit, uncompressed, etc. If available, the ENF frequency is extracted from the audio files, and compared to the database frequency recording for the same date and time. The audio file is an authentic recording and, for the purposes of this experiment, will not be manipulated.

3.4 ENF analysis as a method of detecting audio file manipulations

The purpose of this experiment is to determine if it is possible to detect a manipulation within an audio file by visual examination of the extracted ENF signal of a questioned recording. It is suggested that a manipulation in the audio file is likely to present as an abrupt discontinuity in the phase of the ENF.

The recording that is used for this experiment was manipulated in its original format, and then processed to isolate the ENF component of the signal. The ENF signal is then examined to determine if the manipulation is detectable. This will ascertain if ENF analysis is suitable for use as an authentication method in terms of manipulations without the availability of a reliable database.

The same signal is then also compared to the database recording for the same date and time in order to determine if the manipulation is visible when compared to the ENF database. This experiment thus also investigates the possibility of the temporal validation of a digital audio recording.

3.5 Limitations of ENF analysis as an authentication tool

The various processes of the empirical validation establish the conditions of use and limitations of ENF analysis as an authentication tool, as well as the conditions and infrastructure necessary to record the ENF signal. This provides information regarding the conditions in which the ENF can be recorded (along with the intended audio signal), as well as details relating to the extraction and processing of the audio file. This information assists in the evaluation of the required infrastructure for ENF authentication in South Africa.

The methodologies described above address the research questions posed within this mini-dissertation, and ultimately determines the possibility and limitations of the use of ENF analysis within a South African context.

CHAPTER 4

Results and discussion

Following are the results of the practical validation as stated in Chapter 3 and discussion thereof.

4.1 Recording and extracting the ENF signal

Figure 1 shows an example of the ENF signal extracted from an audio recording recorded at 44.1 KHz, 16-bit, using a Shure SM57 dynamic microphone and an external audio interface.

This indicates that under circumstances, it is possible to record and extract the ENF signal from an audio recording. The equipment that is used to produce the recording must be powered by the electrical network, and not have any electrical components that could affect the ENF signal (such as the circuitry in a condenser microphone). A similar observation can be made from a waterfall spectrogram of the audio file as shown in Figure 2.

The ENF signal is represented as a discontinuity within the audio file at 50 Hz. Using the digital signal-processing methods outlined below, it is possible to isolate this 50 Hz signal, and create a visual representation as frequency as a function of time. This extracted ENF signal can then be used for comparison against a known database of ENF values for further authentication analysis. The following is an outline of the ENF extraction process:

- Application of a 50 Hz bandpass filter on the signal, followed by
- Downsampling of the signal to 150 Hz,
- Computation of the frequency vs time plot through FFT.

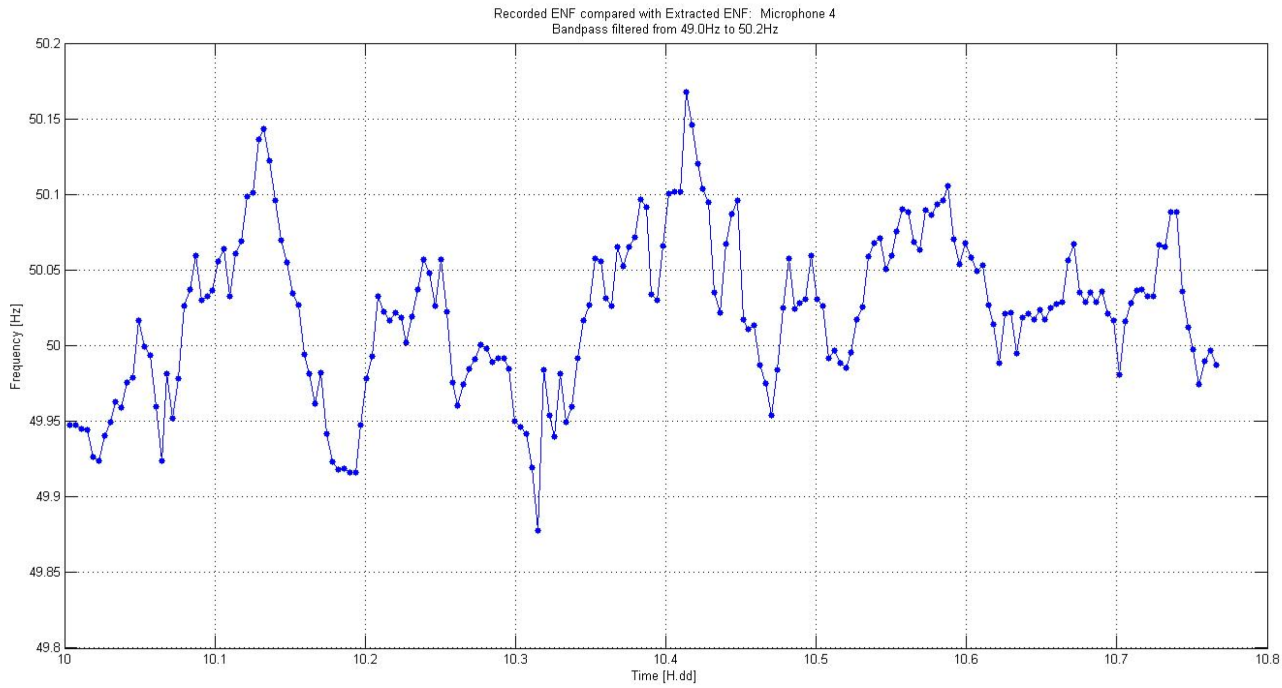

Figure 1 – The frequency vs time trend of the ENF signal isolated from an audio recording utilising the Shure SM57 microphone.

Variables that can influence recording of the ENF signal include the type of microphone used to record the audio, the recording format, the sampling rate, and the bit rate.

4.1.1 Impact of microphone characteristics on the ENF signal

Different microphone technical designs can affect the ability to record the ENF signal, thereby yielding different ENF detection results. This could be attributed to the electronic circuitry in the microphone. It can be expected that differences in the ENF signal detection are primarily a result of the microphone characteristics, since all other parameters are kept constant.

Figure 2 is a waterfall plot of a recording made with a Shure SM57 dynamic microphone. The plot displays a continuous discontinuity at 50 Hz. Extraction yields a clear ENF signal that is considered suitable for ENF analysis.

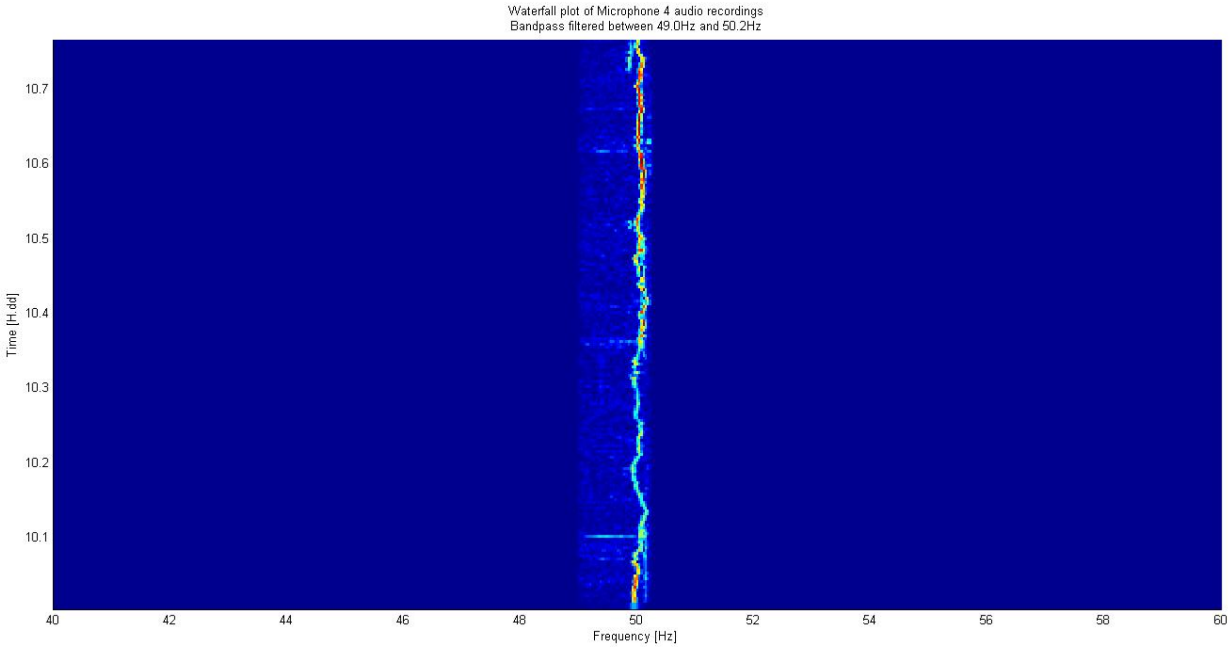

Figure 2 – Waterfall spectrum of the isolated ENF signal extracted from a recording utilising the Shure SM57 microphone.

Figure 3.1 is a waterfall plot of a recording made with a Rode NT2-A, large diaphragm condenser microphone. As seen in Figure 3.1, the condenser microphone did not record the ENF signal and, where present, the ENF is severely distorted (Figure 3.2), and is thus not suitable for ENF analysis. Figure 3.2 compares the extracted ENF signal as produced with the Rode NT2-A condenser microphone, to the ENF signal recorded with the data acquisition unit (that was connected to main electrical supply).

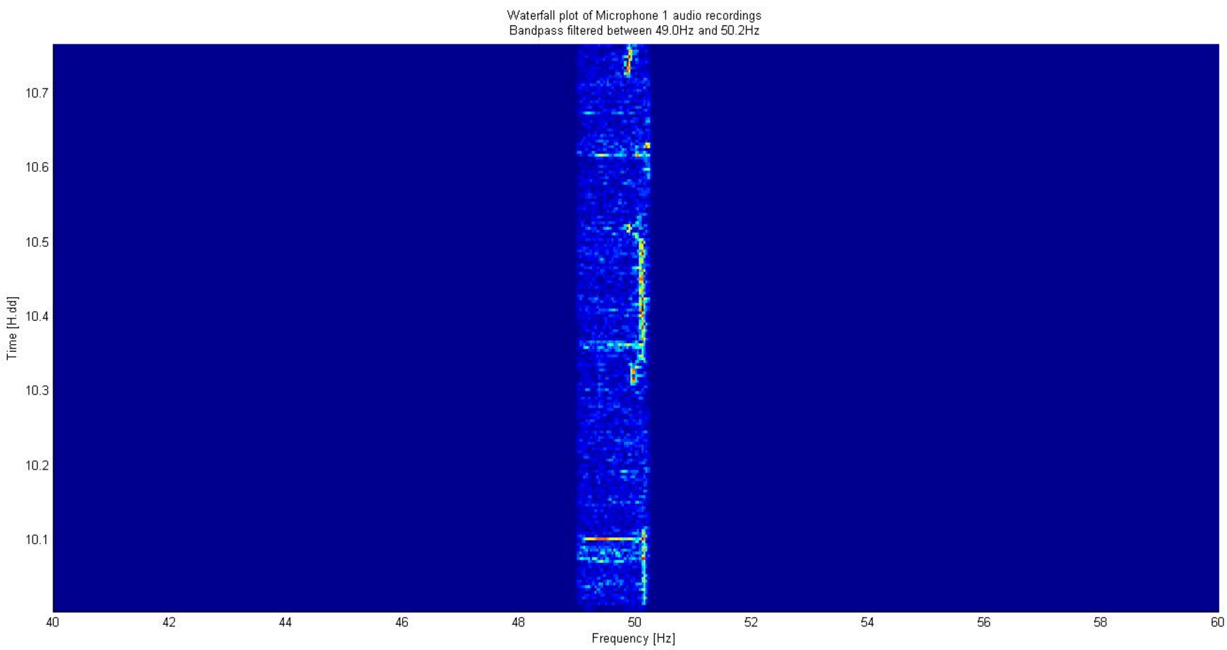

Figure 3.1 – Waterfall spectrum of the isolated ENF signal extracted from a recording utilising the Rode NT2-A microphone.

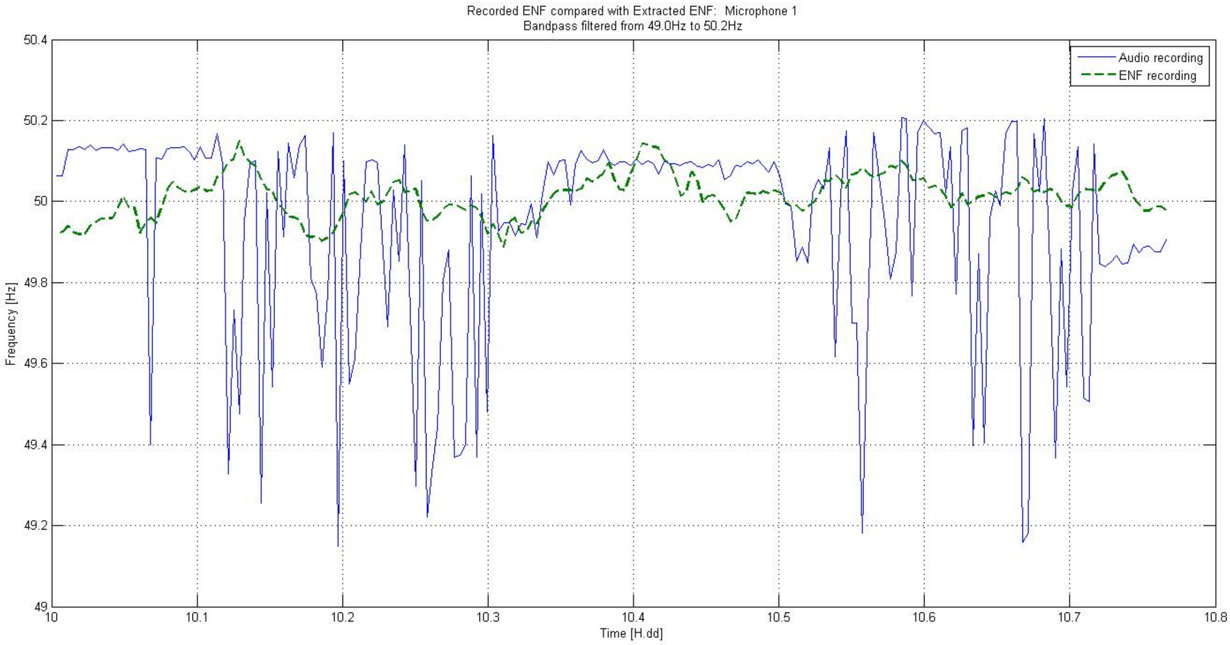

Figure 3.2 – ENF signal from a condenser microphone. The blue signal indicates the extracted ENF from the audio file recorded with the Rode NT2-A, while the green is the ENF recoded with the data acquisition unit from the database.

The recording format, time of production and extraction parameters for both recordings, and the DAW were identical. From the above, it is found that the microphone that yields the more dependable result, and is more suited to the purposes of ENF analysis within this experimentation, is the dynamic microphone (Shure SM57).

4.2 Establishing and maintaining an ENF database

Audio signal authentication is dependant on a reliable database for comparison. This database can be established by recording an audio file to produce an extracted ENF signal, or by means of a data acquisition unit measuring voltage directly from the distribution point.

4.2.1 ENF database via an audio recording

Creating a database by means of recording audio poses the following practical challenges: such as the large data storage requirements, as well as the intensive processing that is required to extract the signal from the audio file. The time required to process the ENF signal is found to be approximately 10 minutes for a 60-minute recording (utilising an Intel i5, 2.5GHz, quad core processor). In addition, common audio recording software does not have the processing capabilities for the requirements of ENF extraction. The software required for the computation of the high-resolution spectrogram or FFT is specialised, costly, and requires specialised computer skills. These issues could jeopardise the integrity of the ENF database.

The need for a continuous signal would require the use of a UPS device to prevent data loss in the case of a power failure. However, since a UPS produces AC power through an electric inverter circuit linked to a DC storage battery, the use of a UPS device would affect the validity and accuracy of the ENF signal. Furthermore, using a DAW, data is not saved or stored intermittently in memory. The recording would need to be stopped and saved manually in a sequenced fashion. Therefore, any interruptions in the power supply would result in data loss.

There are also multiple variables within the audio signal chain that could affect the integrity of the ENF data, such as the properties of the microphone, the external audio interface, the digital audio workstation, and the components of the computer (e.g. fan noise etc). All of these aspects would need to be examined in detail to assess their impact on the ENF signal and consequently the ENF database.

In addition to these challenges, the need for silence in order to avoid interference on the ENF signal makes this method impractical. Due to all to these considerations, for the purposes of the present study, the database was not established by means of audio recording.

4.2.2 ENF database via a data acquisition unit

This is the method generally preferred for recording an ENF database (Kantardjiev, 2011:25). Figure 4 shows the experimental setup used to record the database utilising an electronic data acquisition unit.

Figure 4 – Experimental setup used to record the ENF database via a data acquisition unit.

The following observations were made:

Load shedding and power failures posed significant challenges during this experiment. Over the course of the week of data collection, there were three separate power failures. This resulted in a break in the continuous collection of data, as well as a loss of data. It was also not possible to record the ENF values for the rest of the electrical network during the power failure. These power failures could have been a result of electrical faults within the building, or problems experienced within the network. A stable and continuous power supply is essential when establishing a database, since the database needs to be considered reliable and representative.

Data acquisition unit on-board memory capacity limits the amount of ENF data that can be recorded. Once the memory is full, the unit will stop recording the voltage data. This data then needs to be uploaded to the computer. During the time the data is being uploaded, the unit will not continue recording the voltage data. As such, it is suggested that a second unit be setup in a different location within the same network. The upload time of this unit should be different to that of the first unit to ensure that there is no break in the data collection. An additional benefit of having a second unit would be the security of having a backup collection point in the case of a local power failure or equipment fault.

The processing and storage requirements to extract the ENF signal from a data acquisition unit are much less demanding than that of a recorded audio file. The storage requirements for one hour of data from the acquisition unit are 15.45 MB (when recording 16 bit data at a sampling rate of 150 Hz), while an hour of data from an audio recording is 304.13 MB. This would mean that 48 hours worth of data from the acquisition unit requires 741.75 MB while 48 hours data from an audio recording would require 14.6 GB storage space. There are also fewer points of failure in the signal chain of the data acquisition unit that could affect the integrity of the ENF signal, rendering it a more reliable and stable method of data recording. Results of the experiments indicate that this is a viable option to record continuous ENF values for a database. Figures 5.1-5.6 display the ENF values over the period of 48 hours.

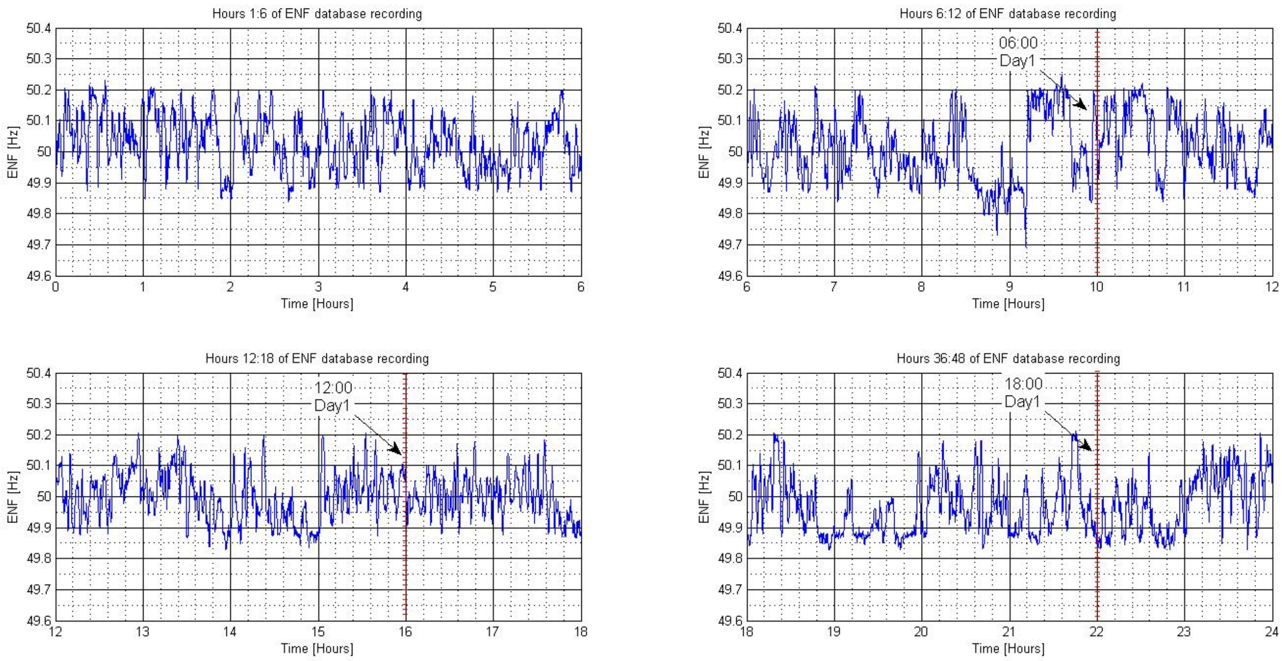

Figure 5.1 – ENF temporal authenticity database recording for hours 0-24 (day 1).

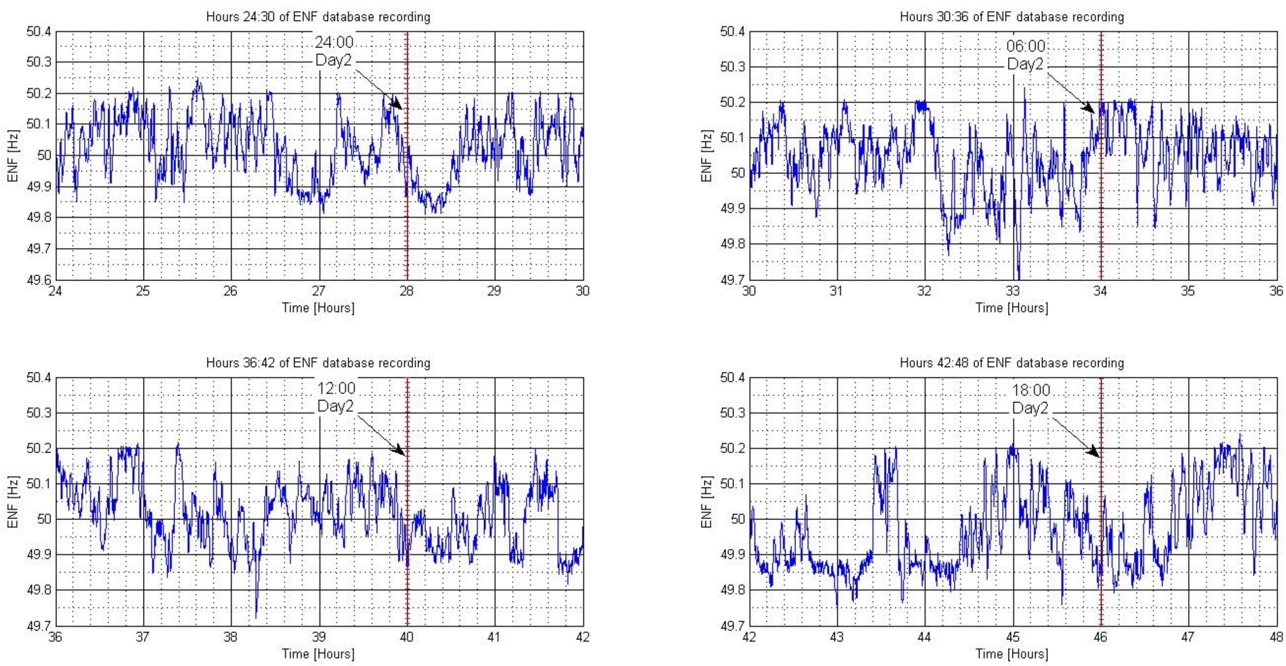

Figure 5.2 – ENF temporal authenticity database recording for hours 24-48 (day 2).

From figures 5.1 and 5.2, it is evident that there is unique patterning in the deviation centred on the 50 Hz signal of the ENF values over time. The only predictable patterns were the dips in the frequency around the daily peak hours. Although the inevitable dip in the frequency as a result of the increase in electricity demand on the power supply is predictable, the exact values of the deviation, or the duration thereof, are not. Figures 5.3 and 5.4 are enlarged images of the frequency deviations around the morning peak period for two consecutive days, while figures 5.5 and 5.6 show the deviations around the evening peak period.

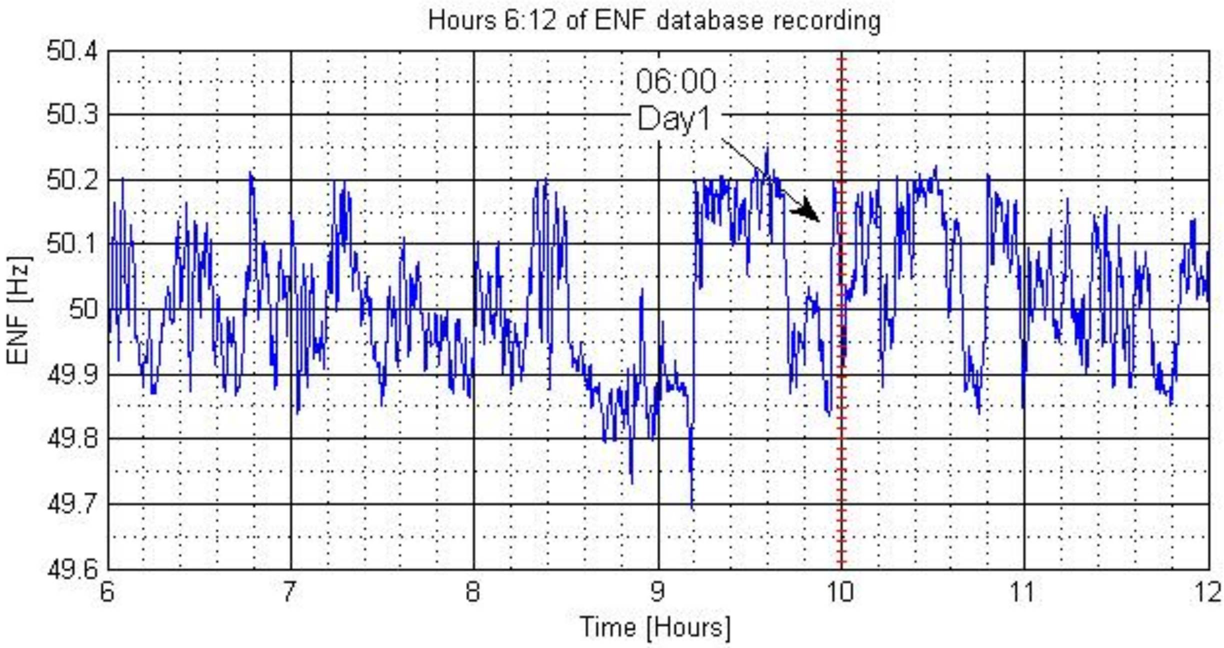

Figure 5.3 – ENF values for the morning peak period for day 1.

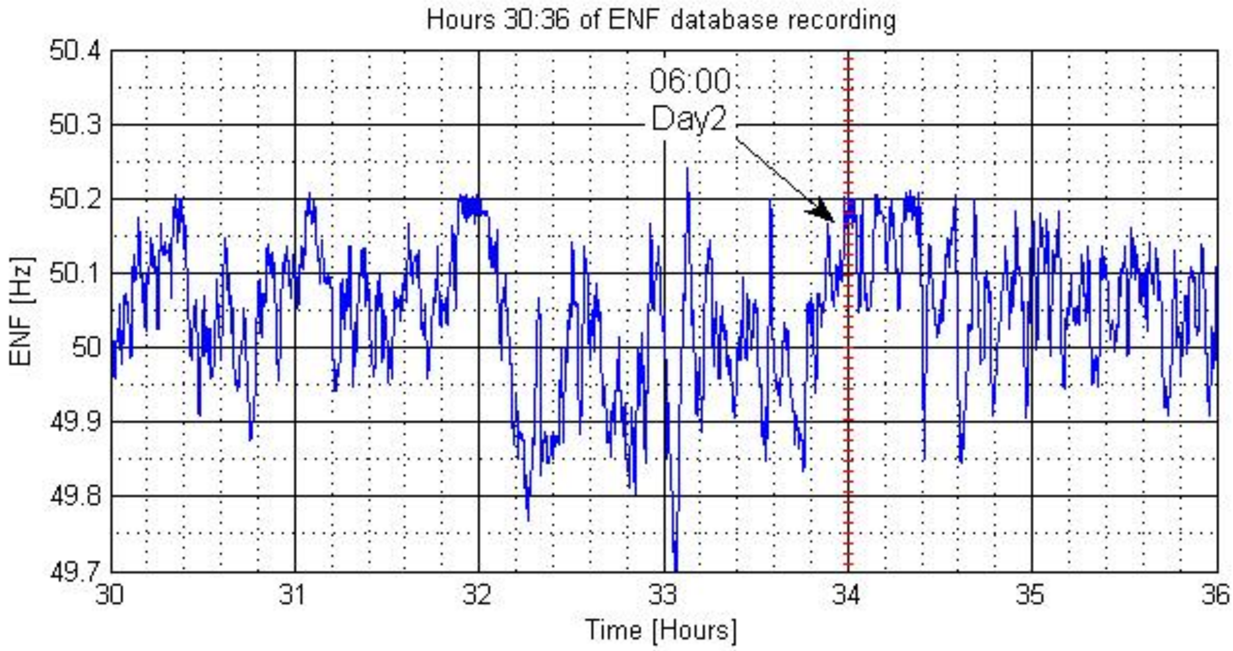

Figure 5.4 – ENF values for the morning peak period for day 2.

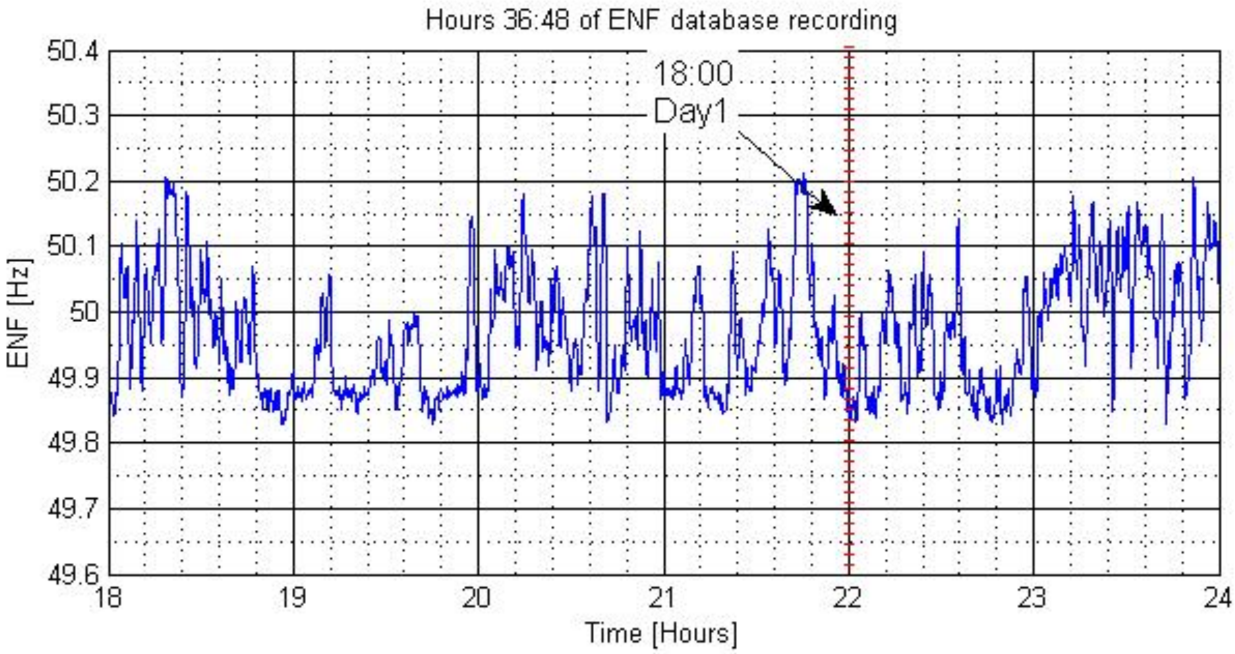

Figure 5.5 – ENF values for the evening peak period for day 1.

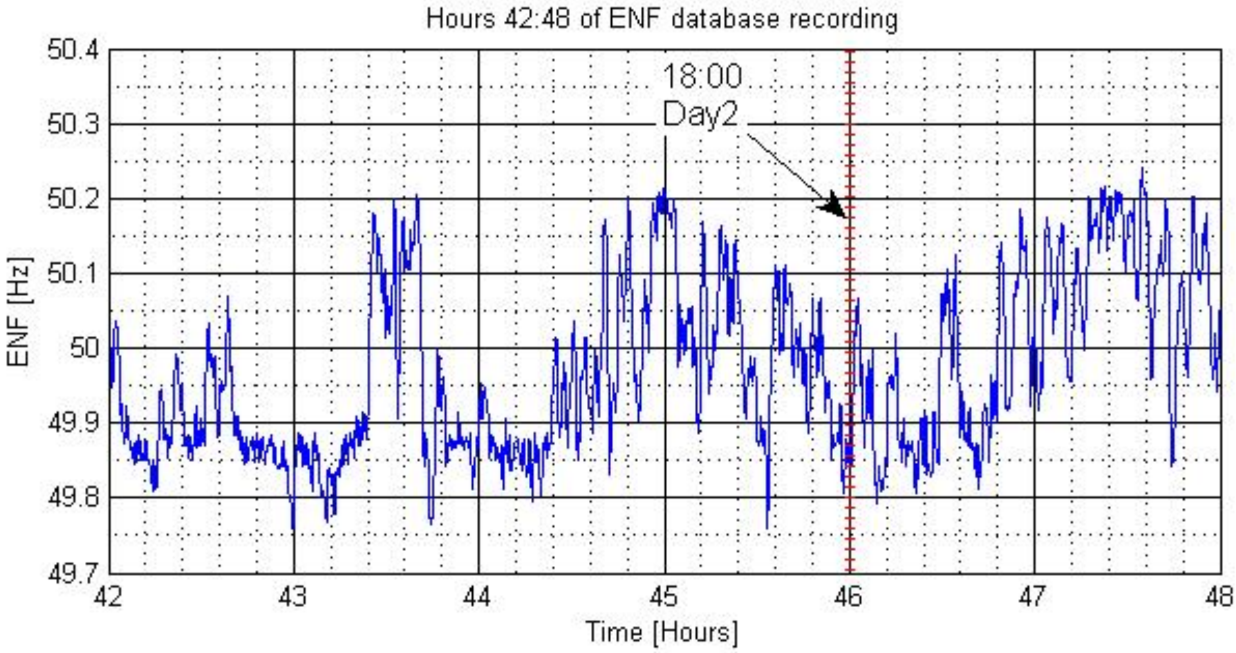

Figure 5.6 – ENF values for the evening peak period for day 2.

The unique nature of the signal makes it a suitable fingerprint for temporal authentication.

4.3 Comparison of extracted ENF signals

The results from the current study suggest that it is possible to obtain temporal validation and examine the integrity of the audio file by comparison of the extracted ENF signals of the audio recording and the reference database. When the two ENF signals are compared to each other, the degree of correlation between the two signals could be sufficient to indicate the authenticity of the audio file. Figure 6 illustrates the comparison of the extracted ENF of the audio file to the ENF of the database for the same date and time.

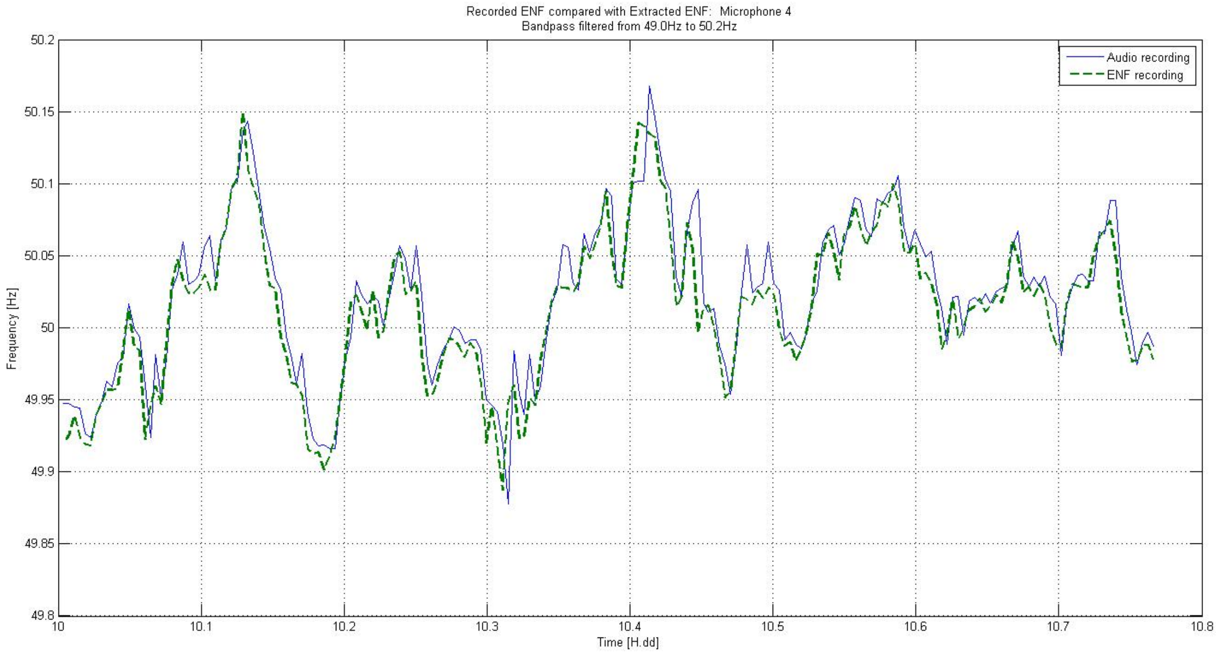

Figure 6 – Comparison of the ENF from an audio file with the ENF database for the same time.

4.4 ENF analysis as an authentication tool

From the previous paragraphs (4.1 to 4.3), it is evident that ENF analysis can be applied for the establishment of the temporal authenticity of an audio file. In order to verify ENF analysis as a technique to detect recorded audio file manipulation, a recorded audio file was edited (a section of 10 minutes was removed from the original file), and the ENF signal extracted. When compared to the ENF of the database, there is a clear discrepancy between the two signals from the point of the manipulation onwards. Figure 7 illustrates the comparison of a manipulated audio file to the ENF database signal. The manipulation within the audio file is clearly visible when the extracted ENF signal is compared with a database of ENF values.

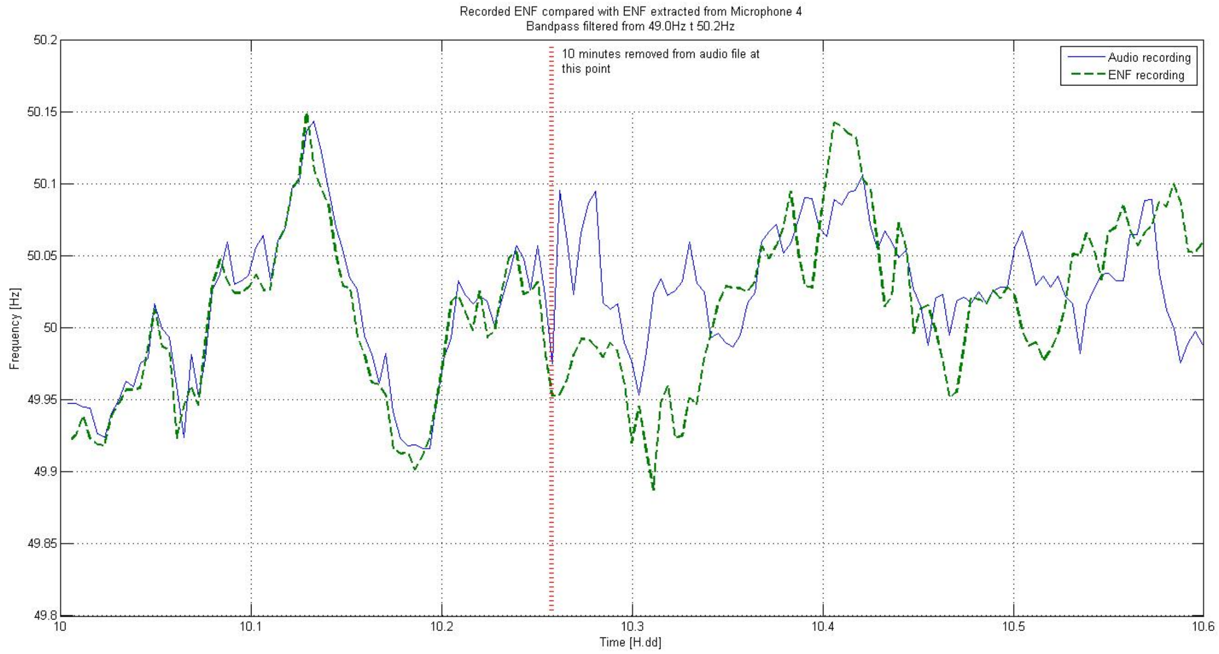

Figure 7 – Comparison of the ENF signal of a manipulated audio file to the ENF database.

The extent to which ENF analysis can be used to identify manipulations within an audio file without a database for comparison is uncertain. Visual inspection of the ENF signal alone does reveal a discontinuity within the audio file that warrants further analysis. However, the examiner would not be able to reach a scientific conclusion regarding the integrity of the file without the availability of the ENF database for comparison. Figure 8 is the ENF signal of the manipulated audio file. Although the point of interest is visible, further analysis or comparison would be necessary to determine the discontinuity.

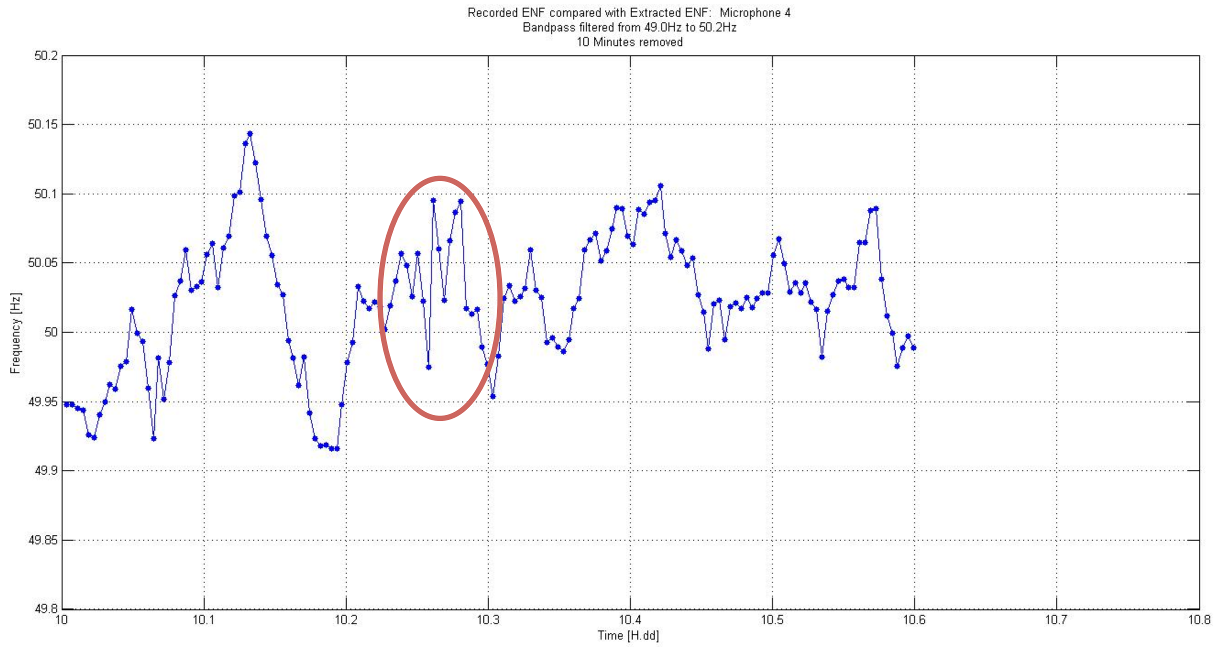

Figure 8 – ENF signal of an edited audio file. A 10-minute segment was removed just before 10.3 (indicated by the circle).

4.5 Limitations of ENF analysis

From Figure 2, it can be seen that the effect of speech on the ENF signal appeared negligible, most likely due to the lack of frequency content around 50 Hz in human speech. However, it is, in principle, possible that the extraction algorithm can misinterpret speech for the ENF. In this instance, it is necessary for the examiner to manually differentiate between the data points. The effect of low-level noise on the ENF signal was not investigated within this practical validation.

It is not possible to extract the ENF signal from a battery-operated recorder. Tests conducted using a Zoom H4N handy recorder in an electromagnetically 'noisy' environment indicate that the ENF was not induced into the audio file.

CHAPTER 5

Conclusion

The results of the practical validation indicate that ENF analysis is a viable option for audio authentication. The electrical supply frequency is suitable for the authentication of digital audio recordings. As from literature, the deviations around the nominal supply frequency of 50 Hz are unique and can be considered random. These random deviations in the supply frequency provide a suitable reference “fingerprint” for temporal validation.

Given specific prerequisites, it is possible to record and extract the ENF signal from an investigated audio recording. The extracted signal can be compared to the extracted ENF database signal for the same time period. The extraction methodologies and parameters suggested by the Best Practises Guidelines needs to be amended to accommodate the wider ENF deviations of South African power supply.

It is evident that the ENF signal is susceptible to interference from variables within the signal chain. Microphone circuit design can influence the accuracy of the ENF recording, to the point where it may be impractical for the purpose of ENF analysis. As such, recordings produced using a condenser microphone are not suitable for ENF analysis. However, recordings produced on a dynamic microphone are well suited for ENF analysis. It is therefore important that all equipment is extensively researched and validated for ENF analysis utilisation.

In order to establish the authenticity of a digital audio file, it is necessary to have a database of ENF values against which an extracted ENF signal from audio files can be compared. Establishing a database of ENF values by utilising a data acquisition unit, and recording an audio file and extracting the ENF values proved feasible. However, the utilisation of an electricity data acquisition unit that measures the voltage directly from the household electricity supply does seem to be more practical. Establishing a database through audio recording and ENF extraction involves many variables within the signal chain that could jeopardise the integrity of the database.

It is possible to detect points of forensic interest by visually inspecting the extracted ENF signal of a recorded audio file. It must be noted that this method is currently not sufficiently developed to yield conclusive results and data. When compared to a reference ENF signal, it is possible to detect manipulations made to an audio file.

The practical validation revealed some specific conditions of use:

- The device used to record the audio must be powered by the electric network,
- The ENF data extraction parameters must be the same for the audio file and the database,
- The microphone must be validated in terms of suitability for recording the ENF signal, and
- The availability of a reliable temporal ENF database is a requirement for authentication.

Results of the practical validation support the hypothesis that ENF is a valid method of authenticating digital audio files. This method can provide supporting information regarding the authenticity of an audio file, and in a forensic scenario, should be used in conjunction with other techniques and methods.

CHAPTER 6

Suggestions for further research

A large number of topics that are directly related to ENF analysis are not addressed within the scope of the present study.

The following are suggested topics for further study:

- Investigate the effect of low-level noise on the ENF
- Determine the effect of a compression algorithm on the ENF signal
- Investigate the applicability of ENF analysis using different file formats
- Determination of the minimum duration of the audio recording for authentication
- Effect of a UPS device on ENF analysis
- Investigation into the efficacy of the various ENF extraction methods
- Investigation into the robustness of an ENF signal – what are the effects of interference on the local network on the stability of the ENF signal
- Investigation into the random nature of the ENF signal
  - i.e. what are the qualities that define randomness and whether or not they apply to the ENF signal
- Refinement of the detection of manipulation in an audio file within the ENF data, without the utilisation of a reference ENF database

SOURCES

Bak, N. (2004). *Completing your thesis: A practical guide.* Pretoria: Van Schaik.

Chang, F. and Huang, H. (2010). *Electrical network frequency as a tool for audio concealment process.* Paper presented at the 2010 Sixth international conference on Intelligent information hiding and multimedia signal processing. (IIH-MSP 2010).

Chang, F. and Huang, H. (2011). *A study on ENF discontinuity detection techniques.* Presented at the 2011 Seventh international conference on Intelligent information hiding and multimedia signal processing. (IHH-MSP 2011).

Department of energy. (www.energy.gov.za/files/electricity_frame.html [retrieved from 4.3.14]).

Electric Network Frequency Discussion Paper. Version 1.2. (2010). Scientific Working Group on Digital Evidence (SWGDE). May.

”Forensics”. Merriam-Webster.com. (2013). http://www.merriam-webster.com [Accessed 2011-07-12].

Garg, R., Varna, L. and Wu, M. (2012). *Modelling and analysis of electric network frequency signal for time stamp verification.* Paper presented at the IEEE International workshop on information forensics and security. December.

Gerazov, B. Kokolanski, Z. Arsov, G. Dimcev, V. (2012). *Tracking of electrical network frequency for the purpose of forensic audio authentication.* IEEE Computer Society.

Glossary of Pro Audio Terms. (ND). Mackie. www.mackie.com. [Accessed 4 March 2013]. USA: Loud Technologies inc.

Grigoras, C. (2003). Digital audio recording analysis, The electric network frequency criterion. *Diamond Cut Productions inc., application notes AN-4.* October.

Grigoras, C. (2007). Applications in ENF criterion in forensic audio, video, computer and telecommunication analysis. *Forensic science international.* Vol. 167, no. 2-3, pp. 136-145. April.

Grigoras, C. (2009). Applications of ENF analysis in forensic authentication of digital audio and video recordings. *Journal of the Audio Engineering Society, Vol. 57, No. 9,* September.

Grigoras, C., Cooper, A. and Michalek, M. (2009). *Forensic speech and audio analysis working group - Best practice guidelines for ENF analysis in forensic authentication of digital evidence.* ENFSI – FSAAWG. June.

Gupta, S., Cho, S., Kuo, C.-C. (2012). *Current developments and future trends in audio authentication.* IEEE Computer Society.

Hofstee, E. (2006). *Constructing a good dissertation: a practical guide to finishing a Master's, MBA or PhD on schedule.* Sandton: EPE.

Hong, G. Yin, Z. (2013). *Research on digital audio authenticity analysis*. Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013). Atlantis Press: Paris.

Huber, E., Runstein, R. (2005). *Modern recording techniques*. 6th Edition. USA: Elsevier.

Huijbregste, M. and Geradts, Z. (2009). *Using the ENF criterion for determining the time of recording of short digital audio recordings.* Proceedings of the 3rd International Workshop on Computational Forensics. pp. 116-124, Heidelberg: Springer.

Izhaki, R. (2008). *Mixing audio, concepts, practices and tools.* Oxford: Focal Press.

Kajstura, M., Trawinska, A. and Hebenstreit, J. (2005). Application of the electrical network frequency (ENF) criterion, A case of a digital recording. *Forensic science international.* Vol. 155. pp. 165-171.

Katz, B. (2007). *Mastering Audio the art and the science.* USA: Focal Press.

Koch, R., Dold, A., Johnson, P., McCurragh, R. and Thenga, T. (2007). *The evolution of regulatory power quality standards in South Africa (1996 to 2006).* Paper presented at the 19th International Conference on Electricity Distribution. May 2007. Paper no. 0894. CIRED, Vienna, Austria.

Koenig, B., Lacey, D. and Killion, S. (2007). Forensic enhancement of digital audio recordings. *Journal of the audio engineering society.* Vol. 55. No. 5. May.

Koenig, B. and Lacey, D. (2009). Forensic authentication of digital audio recordings. *Journal of the Audio Engineering Society.* Vol. 57, No. 9, September.

Koenig, B. and Lacey, D. (2012). An inconclusive digital audio authenticity examination: A unique case. *Journal of Forensic Sciences.* Vol. 57, No. 1. January.

Liu, Y., Yuan, Z., Markham, P., Conners, R. and Liu, Y. (2011). *Wide-area frequency as a criterion for digital audio recording authentication*. Proceedings of the IEEE Power and Energy Society General Meeting. pp 1-7.

Maher, R. (2009). Audio forensic examination. Authenticity, enhancement and interpretation. *IEEE signal processing magazine*. Vol. 26, No. 2. pp 84-94. March.

Michalek, M. (2009). The application of powerline hum in digital recording authenticity analysis. *Problems of forensic sciences 2009, vol. LXXX, pp355-364.*

Mouton, J. (2001). *How to succeed in your master's and doctoral studies. A South African guide and resource book.* Pretoria: Van Schaik Publishers.

Nicolalde, D. and Apolinário, J. (2009). *Evaluating digital audio authenticity with spectral distances and ENF phase change.* Presented at the IEEE International conference on acoustics, speech and signal processing, 2009. April.

National Research Council. (2009). *Strengthening forensic science in the United States: a path forward.* National Academies Press.

Rodriguez, D., Apolinário, J. and Biscainho, L. (2010). Audio authenticity: Detecting ENF discontinuity with high precision phase analysis. *IEEE transaction on information forensics and security.* Vol 5, No. 3. pp 534 – 543. September.

Rumsey, F. (2008). Forensic audio analysis. *Journal of the audio engineering society.* Vol. 56. No. 3. March.

Rumsey, F. (2009). New frontiers in audio forensics. *Journal of the audio engineering society.* Vol. 57. No. 9. September.

Staff Technical Writer. (2006) Audio forensics. *Journal of the audio engineering society.* Vol. 54. No. 1/2. January/February.

Standards South Africa. (2003). *Electricity supply - quality of supply. Part 2: Voltage characteristics, compatibility levels, limits and assessment methods.* 2nd edition, 2003. Pretoria: Standards South Africa. Retrieved June 15, 2011 from http://www.ner.org.za.

The roles of voltage and frequency in the transmission of electricity. TD_0004Rolesrev1-1.pdf. Eskom Fact Sheet. www.eskom.co.za. [Accessed 21 August 2013].

Wang, P. (2007). *How to write a research proposal*. Trinity Western University. <www.meaning.co/archives/art_how_to_write_P_Wang.htm>. [Accessed 2011-08-17].

Watney, M. (2009). Admissibility of electronic evidence in criminal proceedings: An outline of the South African legal position. *Journal of information, law & technology (JILT).* http://go.warwick.ac.uk/jilt/2009_1/watney [Accessed 2013-06-09].

Dissertations consulted for the purposes of this mini-dissertation:

Andersson, F. (2009). *Determining recording time of digital soundrecordings using the ENF criterion*. Thesis. Sweden: Linköpings Universitet.

Cooper, A.J. (2006). *Detection of copies of digital audio recordings for forensic purposes.* Doctoral Thesis. The Open University.

Kantardjiev, A. (2011). *Determining the recording time of digital media by using the electric network frequency.* Masters Dissertation. Sweden: Uppsala Universitet.

Rappaport, D. (2012). *Establishing a standard for digital audio authenticity: A critical analysis of tools, methodologies, and challenges.* Masters Thesis. Colorado: University of Colorado.

Appendix A

MATLAB programming code utilised for ENF extraction:

```
%   ENFExtract routine
%   50Hz ENF Extraction routine
%   2014 - 09- 09
%   Tarryn Denton
%   Rudi Kroch
%   Revision 2

clear all;
clc;

% Toggle to select when processing ENF signals
ENF_signal = 0; % 1-ENF, 2-Audio

% Perform/skip steps in the processing sequence
readfile = 0;
writefile = 0;
down_sample = 0;
fftcalc = 1;
trendplot = 0;

% % Set data directory
dir = ['C:\Tarryn processing\ZOOM H4N\'];
sdir = dir;

%File name, without filetype
fname = ['STE-000'];

% Read data
if (readfile)
    disp('Reading data');

    if (ENF_signal)
        fs = 150;
        load([dir,fname,'.txt']);
        eval(['y = ',fname,'(:,3);']);
        eval(['clear ',fname,';']);
    else
        [y,fs] = wavread([dir,fname,'.wav']);
    end

    % Filter design
    disp('Filtering...');
    N = 20;      % 20th order filter; use 150 for narrowband

    Fc1 = 45.0;   % 45 Hz start                   !!!!!!!!!!!!!
    Fc2 = 55.0;   % 55 Hz cutoff                  !!!!!!!!!!!!!

    d  = fdesign.bandpass('n,fc1,fc2',N,Fc1,Fc2,fs,'linear');
    hd = design(d,'butter');     % Butterworth filter
    y  = filter(hd,y);

    % Save the filtered results in native matlab filetype
    save([sdir,fname,'.mat'],'y','fs');
```

```
end

%Write filtered data back to wav file - for debugging
if (writefile)
    disp('Writing...');
    load([sdir,fname,'.mat']);
    wavwrite(y,fs,[dir,'rw_',fname,'20_100.wav'])
    clear y
end

if(down_sample)
    % Downsample the data
    load([sdir,fname,'.mat']);  % Load filtered data, if not in memory
    disp('Downsampling...');
    fs_final = 150;             % Desired sampling rate
    Nd = fs/fs_final;           % Downsample factor
    y_ds = downsample(y,Nd);

    % Save the filtered and downsampled data in native matlab filetype
    save([sdir,fname,'_ds.mat'],'fs_final','y_ds');
end

% FFT parameters
nfft = 4096;
overlap = 0.5;

% Sequential fft counter intialisation
i_start = 1;
i_end = nfft;

i = 1;

% Calculate the indices to be computed
if (fftcalc)

    % Load filtered downsampled data if not in memory
    load([sdir,fname,'_ds.mat'],'fs_final','y_ds');

    % Perform sequential FFTs with parameters as selected
    disp('Computing FFTs');
    while((i_end < length(y_ds)))

        disp(['Run :',num2str(i)]);

        y = y_ds(i_start:i_end);

        % Create FFT
        H = hanning(nfft);          % Create hanning window
        Y = abs(fft(y.*H,nfft));    % \
        Y = Y(1:end/2);             %   FFT processing
        z = nfft/2;                 % /

        f = (0:z-1)/z*fs_final/2;   % Create frequency vector

        % Find index of lower frequency in range of interest in frequency
        % vector
        f_low = 45; f_low_ind = round(f_low/f(2));

         % Find index of higher frequency in range of interest in frequency
         % vector
        f_high = 55; f_high_ind = round(f_high/f(2));
```

```
        % Find frequency peak in range of interest and perform weighted
        % averaging to obtain ENF
        [A,Fx] = max(Y(f_low_ind:f_high_ind));
        m_ind = [(Fx+f_low_ind-2) (Fx+f_low_ind-1) (Fx+f_low_ind)];
        F = sum(Y(m_ind).*f(m_ind)')/sum(Y(m_ind));

        % Increment sequential FFT counters
        i_start = 1+nfft*overlap*(i-1);
        i_end = i_start + nfft-1;

        % Store FFT of current section, time of the FFT and computed ENF
        F_trend(i,:) = Y;
        ENF(i) = F;
        T(i) = ((i_start+i_end)/2)/fs_final;

        i = i+1;
    end

    % Save FFT matrix, ENF and Time vector
    save([sdir,fname,'_var.mat'],'f','F_trend','ENF','T');
end

% Deprecated, use dedicated graphing function
if (trendplot)
    load([sdir,fname,'_var.mat'],'f','F_trend','ENF','T');
    %T = 10+(T/60)/60;       % Absolute hour plot
    T = T/3600;              % Relative hour plot

    figure(99)
    subplot(2,2,1)
    plot(T,ENF); grid minor
    subplot(2,2,2)
    plot(T,ENF); grid minor
    subplot(2,2,3)
    plot(T,ENF); grid minor
    subplot(2,2,4)
    plot(T,ENF); grid minor
end
```

Appendix B

MATLAB programming code utilised for creating the waterfall and trend plots:

```
%  Graph creator
%  2014 - 10 - 27
%  Tarryn Denton
%  Rudi Kroch

clear all;
clc;

% Enable waterfall plots and ENF trend plots
water = 0;
trend = 1;

% Name and directory of saved ENF recording variables
fname_E = 'TARRYN1_D26_10_14_T09_46_35_R1_B1_1';
dir_E = 'E:\Tarryn Audio Data\ENF recordings\ASCII\';

% Name and directory of saved Audio recording variables
fname = 'STE-000_c1';
dir = 'D:\Tarryn Audio Data\Battery operated\';

%Load ENF recording
load([dir_E,fname_E,'_var.mat']);
ENF_E = ENF;
F_trend_E = F_trend;

%Load audio recording
load([dir,fname,'_var.mat']);

% Time variable
T = (T/3600)*60;     % Relative time
%T = 10+(T/60)/60;   % Absolute time

% Code for ENF trend plot
if (trend)
    offset = 45;        % Correction in time variances between recordings
    figure(offset);

    % Figure parameters
    plot(T,[ENF(1:end)' ENF_E(offset:(offset+length(ENF)-1))'-f(2)])
    %plot(T,[ENF(1:end)'-f(2)])
    axis([0 65 45 55]);
    grid on
    xlabel('Time [Minutes]'); ylabel('Frequency [Hz]');
    legend('Audio recording','ENF recording');
    title('Recorded ENF:  Battery operated recorder, channel1 - Bandpass
45.0Hz to 55.0Hz');
    saveas(gcf,[dir,'Figures\TrendBatChan1'])

end

% Code for waterfall plot
if (water)

    %Figure parameters
    figure(89)
```

```
    h=surf(f,T,F_trend);
    axis([40 60 min(T) max(T)]);
    set(h,'LineStyle','none');
    title('Waterfall plot Battery operated recorder:  Channel2');
    set(gcf,'Renderer','Zbuffer')
    saveas(gcf,[dir,'Figures/BatChan2'])
end
```