The International
Journal of
Speech,
Language
and the Law

IJSLL (PRINT) ISSN 1748-8885
IJSLL (ONLINE) ISSN 1748-8893

Article

An automated approach to the Electric Network Frequency (ENF) criterion: theory and practice

Alan J. Cooper

Abstract

A relatively recent forensic technique developed to help establish the authenticity of audio and video recordings is the Electric Network Frequency (ENF) criterion. This paper confirms the applicability of the ENF criterion for use in mainland UK and describes an automated approach to matching ENF estimates taken from a questioned recording to a database of ENF values. It will be shown that the automated approach has the added advantage of producing statistical data capable of indicating the strength of the match. Consideration has been given to the robust extraction of ENF signals from evidential recordings and the resulting system and signal processing procedures outlined have been used successfully by the Metropolitan Police Forensic Audio Laboratory in London to extract and match ENF data from evidential recordings to an ENF database.

KEYWORDS ENF; ELECTRIC NETWORK FREQUENCY; AUTHENTICITY; EVIDENTIAL RECORDING; DIGITAL AUDIO; INTEGRITY; EDITING; DATE AND TIME

Affiliation
Metropolitan Police, UK
email: alan.j.cooper@met.police.uk

doi : 10.1558/ijsll.v16i2.193

Introduction

Methods for the authenticity examination of analogue recordings are well established (Koenig 1990, Dean 1991, French 1993) and recently introduced magnetic feature visualisation techniques may be applied to both analogue and digital recordings stored on magnetic media (Boss, Gfroerer and Neoustroev 2003, Bouten, Van Rijsbergen and Donkers 2007). However, the signal conditioning principles associated with digital recordings are completely different to that of analogue and additional techniques for assessing the integrity of a suspect recording are required (Cooper 2005, 2008a, Brixen 2007). One such technique is the 'Electric Network Frequency (ENF) criterion' proposed by Grigoras (2005, 2007a, 2007b, 2009).

ENF relates to the frequency of a networked electricity supply transmission system. The ENF criterion is based on estimating the frequency of power line related signals that may have been recorded alongside the wanted audio data as a result of a digital recorder being connected directly to the network supply, or in the case of battery powered recorders, being in the proximity of electro magnetic fields emanating from the network supply.

The ENF is not constant but changes by small amounts in a random fashion over a period of time. Further, the same instantaneous frequency value is experienced throughout the network (Grigoras 2007a). Thus, extracted ENF data from a recording may be compared to a database of ENF information, allowing the date and time of the recording to be ascertained. Additionally, it may be used to establish if the recording has been edited and where in the recording the edits occur. Therefore, the ENF criterion provides a powerful method to assess the evidential integrity of audio data that has been recorded onto audio, video, computer and telecommunications equipment (Grigoras 2005, 2007a, 2007b, 2009, Kajstura, Trawinska and Hebenstreit 2005).

This article confirms the validity of the ENF criterion for use in mainland UK and describes a method to extract ENF data from evidential recordings that allows automatic searching and matching of the data to an ENF database.

Principles of the ENF criterion

Previous studies on the ENF criterion have been largely limited to transmission systems for countries in continental Europe who form an extensive single network controlled by the 'Union for the Co-ordination of Transmission of Electricity' (UCTE). More recently Sanders (2008) examined intra-grid ENF consistency within the three major grids of the United States power system.

The electric power transmission system in England and Wales, which is not part of the UCTE, is run by the National Grid Company (NGC) and connects power stations and substations in a high voltage network and distribution system known as 'The National Grid'. The NGC also operate electricity interconnection systems linking the transmission network in England and Wales to the transmission systems in Scotland and direct current interconnectors with France and Northern Ireland (Figure 1) (Parliamentary Office of Science and Technology 2001, 2007). It should be noted that direct current interconnections allow the ENF to be independent between the coupled transmission systems.

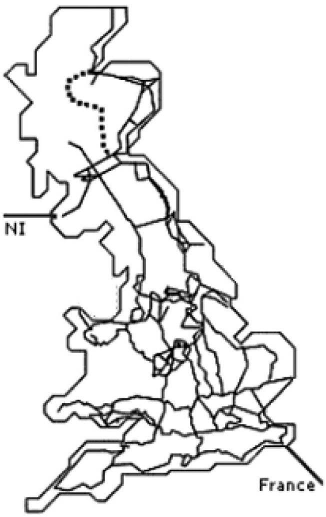

Figure 1: The high voltage power line distribution system ($\geq$275 kV) covering England, Wales and Scotland (Parliamentary Office of Science and Technology 2007).

The majority of power introduced into a network comes from turbines that drive alternating current generators. The turbine's speed of rotation determines the ENF and standards adopted by countries worldwide are based on either 50 Hz or 60 Hz transmission systems; in the UK and Europe the ENF is 50 Hz while in North America it is 60 Hz. An electrical distribution network or grid is organised and power generating facilities distributed in a way that allows the grid operators to cope with wide dynamic changes in supply and demand. Within a network, the generating systems operate in synchronicity, and the ENF will remain constant if the sum of all loads and losses equals the total generation of the network (Sidhu 1999). When there is not enough power available to meet the demand on the grid, the generators all slow down together and the ENF falls. Conversely, if the demand for power drops the generators speed up and the ENF increases. If the average rate of demand on the system differs from the average rate of supply in any given period, the network operators will be presented with a change in frequency for which it must immediately compensate by shedding load for under frequency or shedding generation for over frequency (Sidhu 1999). System frequency will therefore vary around the 50 or 60 Hz target and the network operators have statutory obligations to maintain the frequency within certain limits. For the NGC this is +/- 0.5 Hz, however, in practice it is normally kept within more stringent 'operational limits', which are set at +/- 0.2 Hz (National Grid Company).

The Metropolitan Police Forensic Audio Laboratory has been collating an ENF database in several incarnations over a five-year period. The database located in London has been used to corroborate the date and time of both test and evidential recordings made in places located around England Wales and Scotland. Figure 2 shows a histogram of one month of data taken from the database and is typical of the long-term statistics of the ENF in mainland UK. The ENF has a Gaussian distribution having a mean of 50 Hz and a standard deviation of 60 mHz.

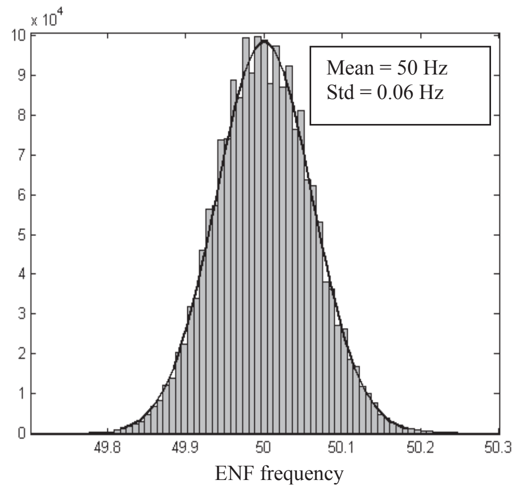

Figure 2: Histogram of ENF data taken from the archive for November 2006. A Gaussian distribution has been overlaid for reference.

ENF components may find their way onto a recording due to poor power supply regulation, earth loops between recording equipment or, more likely, via inductive coupling of ENF currents into high gain recorder circuitry as a result of electromagnetic fields emanating from recorder power supply components such as transformers, or from other nearby mains operated equipment. Under certain circumstances, it is also possible for recordings produced on battery-powered recorders to contain ENF components (Brixen 2008).

The recorded ENF signal may contain harmonics with one or more of the harmonics having a higher power than the fundamental. Therefore, ENF harmonics may also be used for analysis; however, frequencies above the 3rd harmonic are unlikely to be useful due to contamination by lower frequency acoustic signals.

In summary, the synchronicity of the generators produce a uniformity of ENF across any geographic part of the grid and, over a period of time, the dynamic behaviour of the supply and demand provides a unique ENF deviation pattern. The combination of these two factors makes the ENF a powerful forensic tool when an audio recording with a stable time-base has captured the ENF as a by-product of the recording process.

An overview of ENF signal processing

To date there have been three main methods used to extract ENF data from an evidential recording: 'time/frequency domain analysis' based on the spectrogram, 'frequency domain analysis' based on selecting the maximum magnitude of a series of power spectra calculated from consecutive time segments of data, and 'time domain analysis' based on zero crossing measurements of a band-pass filtered signal (Grigoras 2005, 2007a, 2007b). Matching extracted ENF data from a questioned recording to a database of ENF information may be achieved using a visual or automatic comparison. When the questioned recording has a claimed date of origin, verification against an ENF database may be performed by visually comparing ENF waveforms, but when the recording has no claimed date of origin, a visual search could take days or even weeks, making an automated search process attractive.

The methods described in this article are based on a frequency domain approach. Producing a practical method for automatically searching and matching the ENF requires the resulting data from a Fast Fourier Transform (FFT) to be reduced so as to minimise the computational overheads of the search process. The obvious method is to use an estimation algorithm that stores only the peak value of each transform taken over a well-defined bandwidth.

Two problems are encountered when extracting ENF data from evidential recordings using an FFT approach. The first relates to the precision of the measurements made on the ENF signals. This is limited not only by practical considerations, but additionally by the time bandwidth product in Fourier transformation theory known as the 'uncertainty principle'. The uncertainty principle states that you can not obtain arbitrarily high resolution in both the time and frequency domains simultaneously, making low frequency signals that vary with time very difficult to estimate (Czyzewski et al. 2007). The second problem being that in general the recorded ENF signal energy is usually small, producing frequency estimates that are susceptible to error due to noise. A number of signal processing techniques have been proposed to help overcome the limitations of the Fourier transform uncertainty principle, including those based on parametric frequency estimators (Czyzewski et al. 2007), zero padding/interpolation schemes (Abe and Smith 2004) and signal derivatives (Desainte-Catherine and Marchand 2000).

Two main signal processing procedures will be discussed. The first relates to an ENF signal estimation process that is used for both the extraction of data directly from the electricity network used for archiving purposes and from the evidential recordings under examination. The method incorporates an overlapping Short Time Fourier Transform (STFT) combined with a peak interpolation scheme (Abe and Smith 2004), allowing adequate time and frequency resolution to be obtained for minimal computational overheads. The second process describes an automated search algorithm that matches the extracted data from the evidential recording to the archive of ENF values forming the database and is based on a simple Mean Squared Error (MSE) search. Both the extraction and matching algorithms have been developed using MathWorks MATLAB.

Producing an archive of ENF estimates directly from the network

A real time ENF archiving system suitable for automated matching was developed by the author and used as the laboratories primary database until November 2008. For simplicity the ENF for the database was estimated using a commercial off-the-shelf FFT analyser capable of peak picking frequencies over a user defined frequency window. However, as previously stated, the FFT has inherent time and frequency resolution trade-offs (Czyzewski et al. 2007) and in order to obtain acceptable time and frequency resolution for the ENF estimates, a novel process was applied which increased the analysis bandwidth for a directly proportional reduction in the analysis time window. This was achieved by making the original ENF sine wave signal non-linear and analysing a resulting higher order harmonic component (100th). Compared to the fundamental ENF, the harmonic will have a directly proportional increase in bandwidth, and for a given frequency resolution will require a directly proportional reduction in FFT size and therefore a directly proportional reduction in the amount of input data required (Cooper 2008b). However, although the system works well, the off the shelf FFT solution inhibits further development due to a limited set of user accessible parameters.

The author has recently developed an improved real time ENF signal process suitable for archiving purposes, which interfaces to a standalone graphical user interface running under a widows XP environment and developed using Matlab. The system is designed to run continuously, automatically archiving ENF values and associated date and time information to two independent storage locations at regular intervals. Figure 3 shows the GUI for the archiving system. The GUI primarily displays the past hour of ENF information along with the current ENF value and associated date and time.

The ENF signal is digitised using a step-down mains transformer connected directly to the electricity network with its output fed to a high quality 16 bit PC soundcard. After appropriate software interfacing with the soundcard, the ENF is estimated using FFT and peak interpolation procedures to be described in the next section. With consideration to the ENF rate of change of frequency, ENF to noise ratios and signal processing efficiencies, the ENF estimates are produced every 1.5 seconds at a spectral resolution circa 0.7 mHz. Every ENF estimate is stored to a data file along with the associated date and time. Each data file is automatically archived after 10 days, thus allowing automated searches to be carried out on easily manageable chunks of data.

Irrespective of the signal processes applied, the quality of ENF estimates will be limited by the quality of the soundcard or analogue to digital converter. The key parameters of interest are the accuracy and stability of the clock used to derive the sampling rate of the conversion process. It has been found in practice that non-professional soundcards may have a significant offset from the nominal sampling rate, resulting in a proportional frequency offset of the ENF estimate. Poor quality soundcard sampling rates may also suffer from drift due to changing temperature. It is therefore important that high quality professional soundcards/converters are used for ENF database management. Additionally, synchronising the soundcard to an external high quality clock source may be advantageous.

From an evidential perspective confidence in the date and time of the ENF estimates are a requirement. Over a period of time the internal date/time of the PC will drift requiring regular manual adjustment. Synchronisation to an external date/time reference such as may be derived from a network or internet based time-server may provide an automated solution. The laboratories ENF archiving system utilises the MSF 60 kHz radio transmission maintained by the National Physical Laboratory. The radio signal carries highly accurate date and time information across the UK derived from an atomic clock source. This form of date/time synchronisation allows the ENF archiving PC to remain isolated from external networks.

In order to safeguard the archiving PC from crashes due to electricity network power outages, the PC system power comes via an uninterruptible power supply (UPS). A complete ENF archiving system as described is shown in Figure 4.

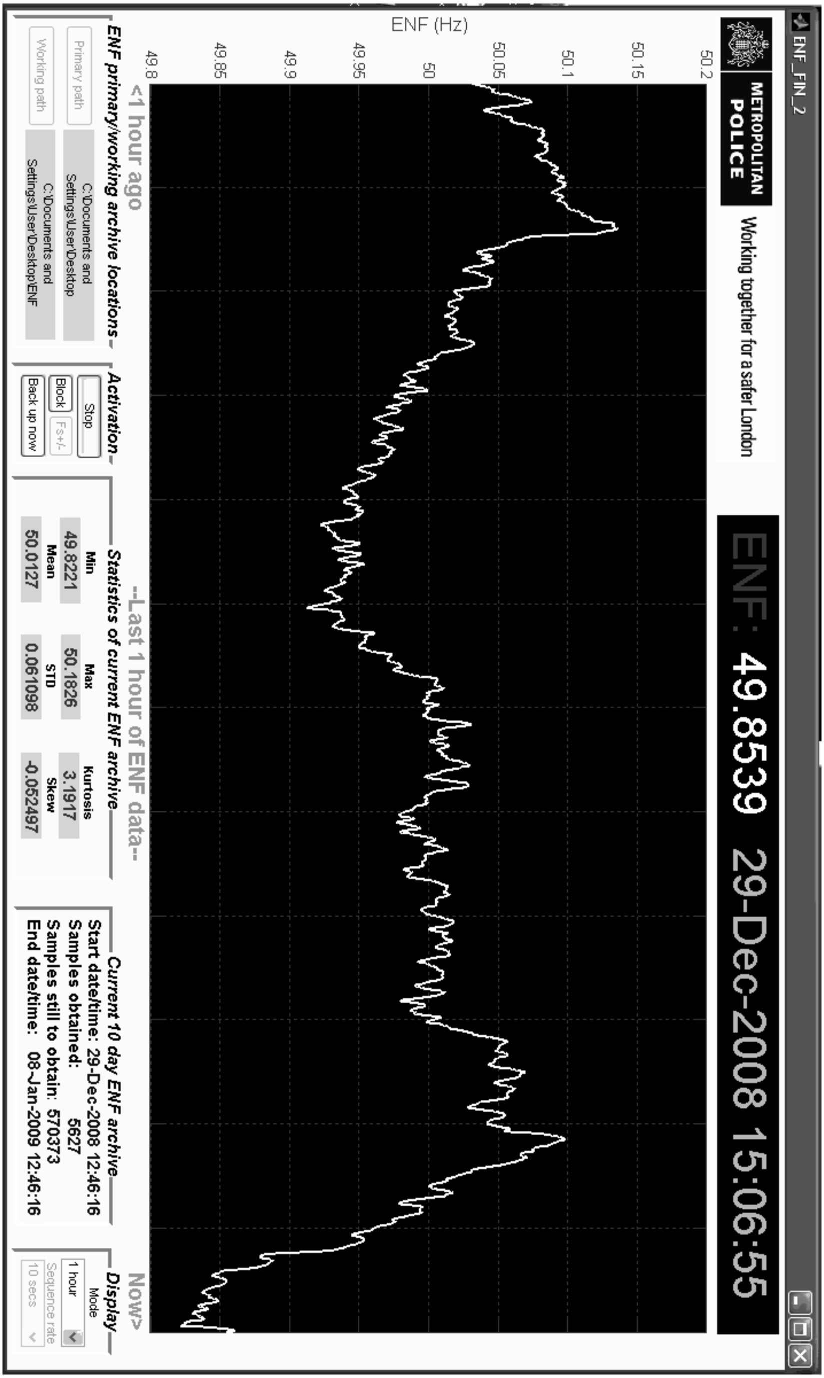

Figure 3: Example screenshot of the ENF archiving system GUI.

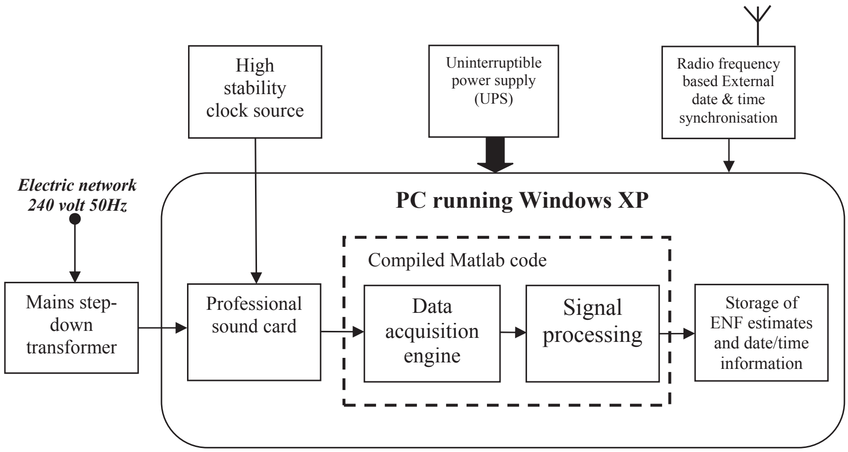

Figure 4: ENF archiving system.

Extraction of ENF data from forensic digital audio and video recordings

The problem of extracting the ENF data from an evidential recording may be defined as follows: track and estimate at regular time intervals a single sinusoidal component having a variable rate of change of frequency, a finite SNR and a relatively narrow bandwidth.

In the presence of interference and low SNR's, a robust estimation of ENF is required. A further requirement is that the extracted data must be in a form that is compatible with the archived data to allow a simple and efficient matching process to be used; and with this in mind the signal processes to be described are the same as used to estimate the ENF for archiving purposes.

From network operational practices as described by the NGC, the total bandwidth of interest for the ENF may be defined over the range 49.5 Hz to 50.5 Hz. The SNR is determined by the induced level of ENF into the recording system and the relative levels of electronic and acoustic noise residing over the ENF bandwidth of interest. The method chosen to track the dynamic behaviour of the ENF is based on the Short Time Fourier Transform (STFT) (Allen 1978). Peak magnitude estimation is achieved using a quadratic interpolation and mild zero padding scheme as described by Abe and Smith (2004). The process leads to an approximately maximum likelihood estimator for moderate signal to noise ratios.

Short Time Fourier Transform (STFT)

The STFT may be derived by splitting the input signal x consisting of $n = 0,1 \cdots N-1$ data samples into M frames:

$$x_m(n) = x(n - mL) \qquad \text{eq 1}$$

Where x_m is the m^{th} frame of the input signal and L is the number of samples advanced between each consecutive frame known as the hop size. Each frame is then multiplied by a spectral analysis weighting window w producing:

$$\tilde{x}_m(n) = x_m(n) \cdot w(n) \qquad \text{eq 2}$$

$\tilde{x}_m(n)$ is then extended by zeros using a factor of b to produce a zero-padded windowed frame $\tilde{x}'_m(n)$. Converting each frame to the frequency domain using a length P FFT produces the STFT at frame m:

$$X_m(k) = \sum_{n=0}^{P-1} \tilde{x}'_m(n) e^{-\frac{j2\pi k f_s}{P}} \qquad \text{eq 3}$$

Where f_s is the sampling rate in Hz and k is the k^{th} frequency bin. The hop size L is set to match the ENF sampling time interval of the archiving database (1.5 seconds) and the FFT transform size P is variable and dependent on a user defined sample interval multiplication factor D, to be described:

$$P = L \cdot f_s \cdot b \cdot D \qquad \text{eq 4}$$

The STFT segmentation process is described by Figure 5.

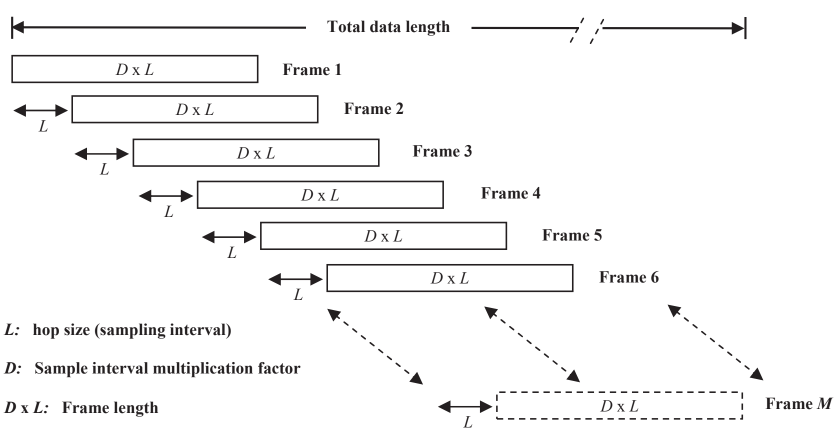

Figure 5: Segmentation and overlap scheme used for the STFT.

Each STFT frame is analysed to find the prominent local maximum or peak in the magnitude spectrum corresponding to the ENF. The ENF peak value in any frame is unlikely to coincide with the exact frequency position of an FFT transform point making for imprecise frequency estimates. One technique that may be deployed is that of zero padding, where zeros are appended to the time domain data prior to the application of an FFT. One advantage of zero padding in the time domain is that it increases the FFT size making each FFT bin bandwidth (f_s / P) proportionally narrower, producing a more densely sampled spectrum providing accurate interpolation in the frequency domain and therefore accurate ENF estimation. However, to gain reasonable accuracy the zero padding factors have to be very large, resulting in very large FFT sizes. This has obvious implications for signal processing efficiency, where time scales are increased according to $P \cdot \log(P)$.

Quadratically Interpolated FFT (QIFFT)

In order to overcome the computational limitations of high zero padding factors, a quadratic interpolation of the ENF sinusoidal spectral peak has been used in conjunction with a mild zero padding factor as described by Abe and Smith (2004). The advantage of interpolation is that it allows a high resolution ENF estimate to be made using a relatively low FFT size and is therefore computationally efficient. The procedure is straightforward; compute the log power spectrum of each STFT frame using a zero padding factor of 4, and then apply a quadratically interpolated FFT (QIFFT) to each frame as follows:

a) Select the FFT bin β having maximum magnitude over the spectral bandwidth of interest.

b) Select the adjacent FFT bins $\beta - 1$ and $\beta + 1$ either side of the peak

c) Fit a second order (quadratic) model to the 3 data points.

d) The peak value (p) of the QIFFT is the estimated peak value of the quadratic model.

From eq 3, the amplitudes of the three frequency points expressed in dB from each frame as defined above are given by:

$$\alpha = 20\log_{10}\left|X_m(k_{\beta-1})\right| \qquad \text{eq 5}$$

$$\beta = 20\log_{10}\left|X_m(k_{\beta})\right| \qquad \text{eq 6}$$

$$\lambda = 20\log_{10}\left|X_m(k_{\beta+1})\right| \qquad \text{eq 7}$$

Solving for the peak location p of the quadratic model using α, β and λ (Smith and Serra 1987):

$$p = \frac{1}{2} \cdot \frac{\alpha - \lambda}{\alpha - 2\beta + \lambda} \qquad \text{eq 8}$$

Estimation error bias inherent in QIFFT is the difference between the true peak value and the peak value of the fitted quadratic model. The bias is reduced to acceptable levels by the application of the mild zero padding factor as described (Abe and Smith 2004). With the parameters discussed, an ENF spectral resolution of circa 0.7 mHz is obtained. Figure 6 shows a parabola fitted to α, β and λ with the resulting interpolated frequency estimate.

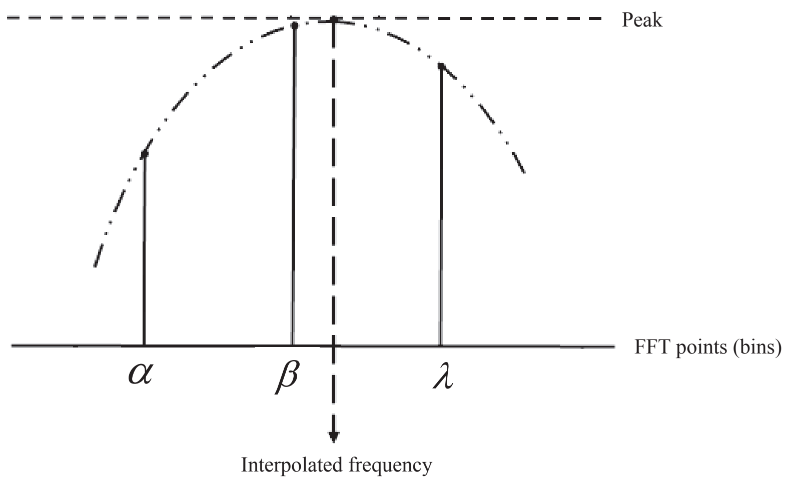

Figure 6: Interpolation using the coarse FFT peak value and its adjacent values.

The overall extraction process

To improve the efficiency of the processing, the initial sampled audio data is decimated to between 120 Hz and 420Hz dependent on the ENF frequency of interest i.e. 50, 100, 150 or 200 Hz. This is followed by a band-pass filter with a bandwidth set to the ENF frequency region of interest of 49.5 Hz to 50.5 Hz for the fundamental 50 Hz component, suitably shifted and scaled for a harmonic component. The time domain signal is then split into overlapping frames and each frame processed as previously described. The required FFT size is the product of the frame length in seconds, the post decimation sampling rate and the zero padding factor. The overall extraction process is shown in Figure 7.

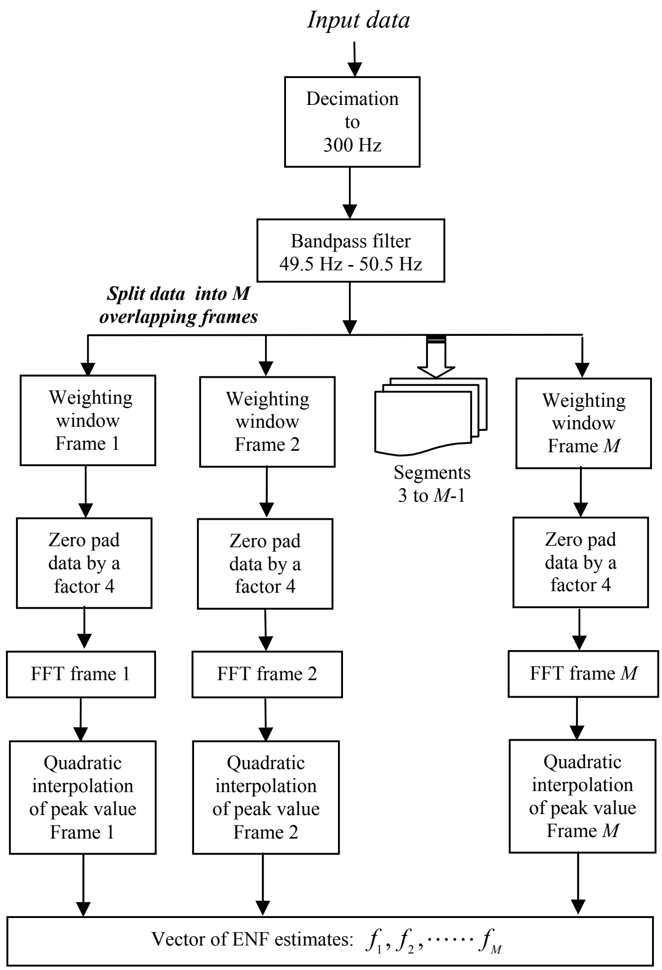

Figure 7: Overall ENF signal processing procedure.

Noise robustness

This section discusses the practical problem of estimating the peak value of a single slowly varying sinusoid from a finite number of noisy discrete time observations. Rife and Boorstyn (1974) established that when using rectangular weighting windows and zero padding factors approaching infinity, an exact spectral peak estimator is the maximum likelihood estimator for a single sinusoid in additive white Gaussian noise. The maximum likelihood estimator will asymptotically achieve the Cramer-Rao lower bound on the estimator variance i.e. the lower bound for the variance of an unbiased estimator (Papoulis and Pillai 2002). Abe and Smith (2004) point out that since the QIFFT is an approximate spectral peak estimator it qualifies as an approximate maximum likelihood estimator under the above conditions.

The application of the band-pass filter provides frequency selectivity, confining the data analysis to the ENF region of interest only. The band-pass filter, which is applied before the signal is segmented, will also eliminate spectral leakage components produced from signals outside the region of interest masking the ENF. This allows a rectangular weighting window to be deployed, leading to greater noise immunity, due to it having the narrowest mainlobe in the frequency domain of all weighting windows (Harris 1978).

Noise reduction using overlapping analysis frames

Stochastic noise is also a major problem for ENF extraction as the level of induced ENF is often very small. The FFT can be very effective in picking out periodic components of a signal, even when it is affected by relatively high noise levels. A P point FFT may be considered as $P/2$ contiguous band-pass filters, with the bandwidth of each filter being dependent on the sample rate and the number of points used in the FFT (f_s / P). Considering the noise to be white a signal/noise improvement can be achieved by spreading the noise over P filters producing a noise power reduction φ of:

$$\varphi = 10\log_{10}\left(\frac{1}{P}\right) \qquad \text{eq 9}$$

Thus, within the FFT frequency bin containing a sinusoid, the SNR will be improved by 3 dB for every doubling of FFT size. However, doubling the FFT size requires a doubling of data size and therefore a trade-off between SNR and time resolution between frames.

For typical evidential recordings, the SNR's for ENF estimation needs a large frame size requiring many seconds of data for each frame. In the overlapping frame system previously described (Figure 5), the parameter D (eq 4) is used to increase the transform/frame size in multiples of ENF sample interval. This process exploits the noise suppression properties of the FFT, allowing a trade-off between noise performance and ENF envelope resolution.

An example of noise performance verses frame length is shown in Figure 8 where a noisy ENF signal has been extracted from a 6 minute long section of recording. Fig 8a shows an extraction using a frame length set to 1 ENF sample interval (D=1), Figure 8b shows an extraction of the same data using a frame length set to 5 ENF sample intervals (D=5) and Figure 8c shows the same data extracted using a frame length set to 15 ENF sample intervals (D=15). It can be seen that increasing the frame length reduces the effects of the noise. It should be noted that ENF signals estimated directly from the electricity network have very good SNR's and therefore no frame overlap is required.

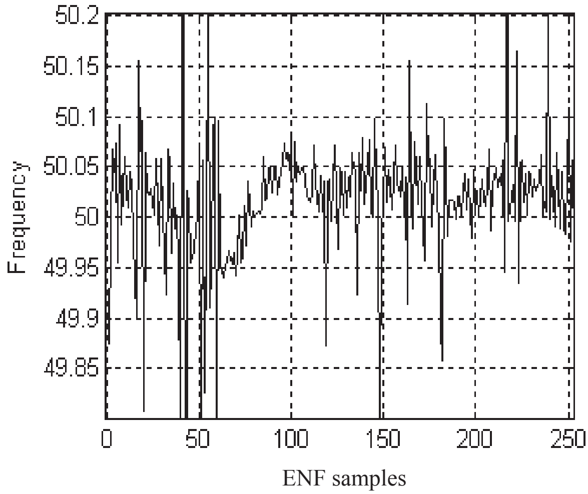

Figure 8a: Frame length set to 1 sample interval.

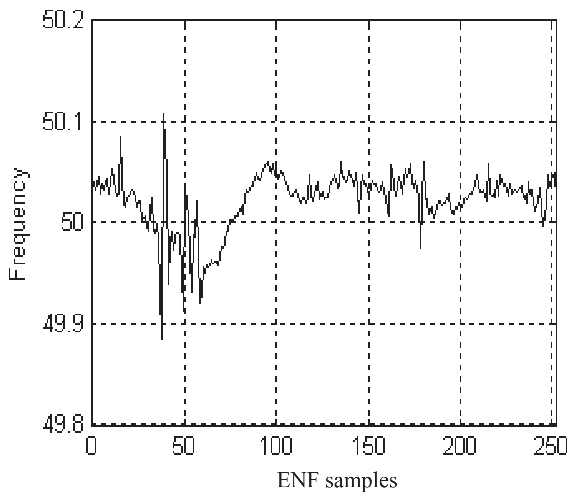

Figure 8b: Frame length set to 5 sample intervals.

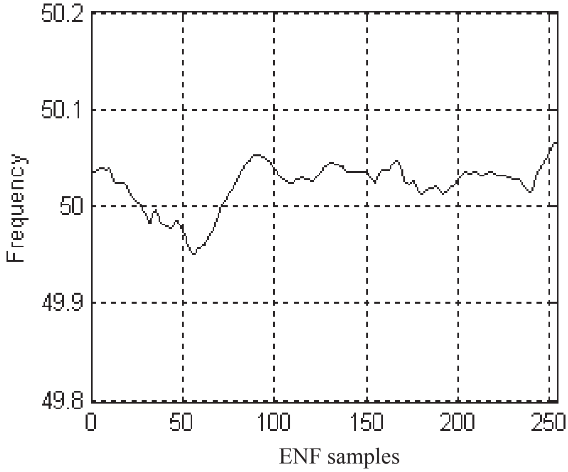

Figure 8c: Frame length set to 15 sample intervals.

Noise reduction using Threshold Dependent Median Filtering (TDMF)

When estimating single frequencies in noise, decreasing SNR's result in increasing peak estimation errors and there is usually a point at which the error rises very rapidly (Rife and Boorstyn 1974). This exhibits itself as spurious spikes within the ENF pattern as demonstrated by Figure 9 which shows an extracted noisy ENF signal taken from a DAT surveillance recording.

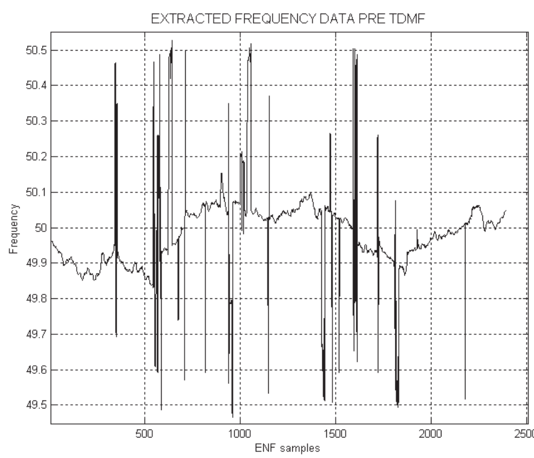

Figure 9: Spurious spikes in ENF pattern due to poor ENF to noise ratio.

These spurious spikes can be reduced by increasing the frame length as previously described. However, a further ENF noise reduction technique developed by the author is 'Threshold Dependent Median Filtering' (TDMF). A waveform containing large numbers of transients, spikes or signal bursts can be very effectively cleaned using median filtering. The concept of median filtering is simple: take a window of data values, sort the values into order of magnitude and then select the value at the centre of the sorted list, move the window on one data sample and repeat the process. The filter may be described by eq 10, where the output is the median value of the input samples taken over a window of length $N=2j+1$. Clearly, any outlier values representing the spikes in the data would be removed by the process. The longer the median filter length, the more effective it may be at removing spikes from the data. However, the median filter is a form of low-pass filtering, increasing the filter length reduces the effective cut off frequency, and this may remove significant detail from the ENF signal not affected by noise.

$$\chi_n = med[x_{n-j}, \ldots, x_n, \ldots, x_{n+j}] \qquad \text{(eq 10)}$$

TDMF has been developed to remove noisy spikes while leaving good quality data unaffected, its operation is shown in Figure 10 and described as follows:

1. A relatively high n^{th} order median filter is applied to the data set. As discussed, the data is effectively low-pass filtered, removing both the impulse noise and the fine detail, leaving the trends in the original data. In practice a median filter having a length of $n=50$ is found to adequately model the trends.
2. Subtracting the estimated trend from the original signal produces a 'detrended' version and effectively leaves the impulse noise and the finer detail of the wanted signal.
3. Based on the hypotheses that the impulsive noise energy is greater than the energy in the signal detail, a threshold point on the detrended signal can be identified that differentiates between the noise and the signal detail. As the detrended signal is bipolar, the threshold value is also bipolar (having an identical positive and negative value).
4. For each sample in the detrended ENF signal, a comparison to the bipolar threshold value is made. If the detrended value is outside of the threshold limits the sample is considered noise and the original sample is replaced by the value from the filtered signal, if the value is within the threshold limits it is considered signal and the original sample is left intact.

With the TDMF method, only noise-contaminated ENF sample estimates (exhibiting themselves as spikes in the waveform) are replaced with the median filtered versions, leaving good quality ENF estimates intact and thus avoiding the loss of higher frequency ENF waveform details as a result of indiscriminate low-pass filter action.

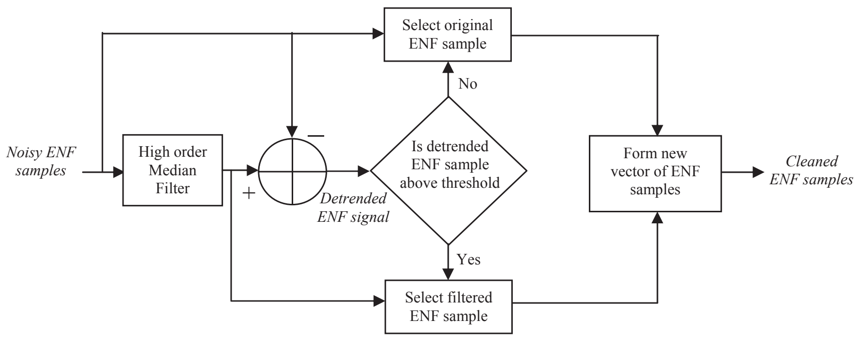

Figure 10: Principle of the ENF Threshold Dependent Median Filter.

The initial bipolar noise threshold value is preset to ±3 standard deviations of a detrended clean ENF signal (circa ±0.02 Hz). To obtain an optimal bipolar threshold value, a 'detrended threshold confidence graph' may be used. This graph represents the detrended ENF signal overlaid by a chosen bipolar threshold value. The bipolar threshold value may therefore be chosen, based on the visible discrimination between the likely signal components and the impulsive noise, as shown in Figure 11. By applying the TDMF process to the noisy ENF data identified in Figure 9, and using the detrended data trace to select a near optimum threshold value of ±0.03 Hz (Figure 11), the resulting filtered data becomes impulse free as indicated by Figure 12.

In practice both of the noise reduction techniques described (overlapping frames and TDMF) may be used for both visual matching and automated matching of ENF extracted data to ENF archived data.

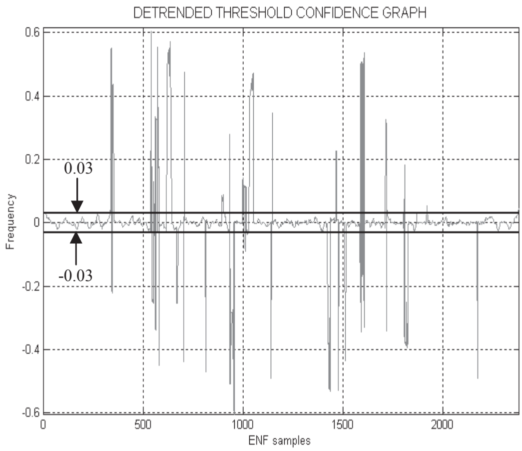

Figure 11: Graph showing the detrended noisy ENF signal with a bipolar threshold value chosen to discriminate between signal and impulse noise.

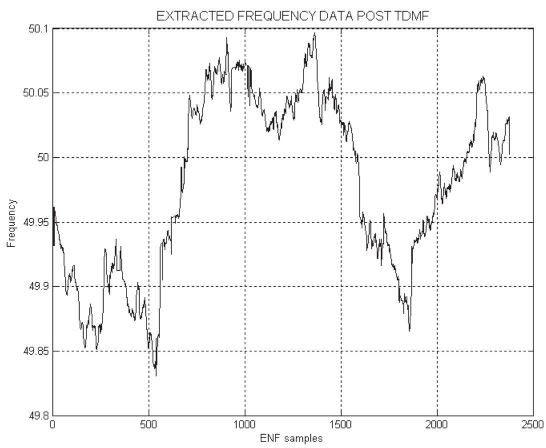

Figure 12: Impulsive ENF estimate after processing by TDMF.

Automated matching of ENF data

An algorithm is required that searches for the extracted ENF pattern in the archived ENF data. The process to be described involves overlaying two length N vectors a and t and computing a metric that determines the degree to which the two vectors match. The metric chosen for this matching process is based on the mean squared error (MSE). Other metrics have been considered, including cross-correlation (Huijbregtse and Geradts 2009). However, it was found in practice that cross correlation had higher processing time overheads compared to the MSE approach, leading to significant time penalties when searching long ENF databases.

The extracted data is overlaid at the start of the archive file and the MSE between the extracted data and the part of the archive that has been overlaid is calculated, the result is then stored. The extracted data is moved one sample forward and the process is repeated until the extracted data has slid across the entire archive file. A vector of 'error' values is therefore formed and the minimum value is tagged as the best match. This error value directly indicates the start position of the match in the archive and therefore the date and time. The error ε is given by the logarithm of the MSE:

$$\varepsilon = \log\left(\frac{1}{N}\sum_{i=0}^{i=N}(a_i - t_i)^2\right) \qquad \text{eq 11}$$

Where a_i is the i^{th} element of the overlapped archive file and t_i is the i^{th} element of the extracted file, N is the number of samples or elements of the extracted file. The logarithm provides better visual discrimination when the errors are plotted graphically and produces less skew in the overall error distribution aiding statistical analysis.

It should be noted, that in order to make the search algorithm immune to ENF frequency offsets resulting from a bias in the original sampling rate of the recorder used to make the recording under analysis, vectors a and t should have there mean values subtracted prior to the application of eq 11.

An example showing the results of an automated matching process using the techniques described is shown in Figure 13. The extracted data is from a 70 minute recording produced in Glasgow Scotland UK and the archive produced in London England UK, a distance of 420 miles (676 km), the extracted data has been offset by 0.1 Hz to aid visual comparison.

Statistically the strength of the match may be ascertained by examining the minimum MSE value in relation to all the other MSE values obtained during the search of the archive. It is found that for relatively long recordings > 30 minutes, the error distribution may be close to Gaussian. Figure 14 shows the standard errors and Figure 15 shows a histogram of the standard errors with a Gaussian distribution overlaid for comparison purposes. It can be seen that the minimum error is >6 standard errors below the mean, indicating that the match has almost certainly not resulted by chance.

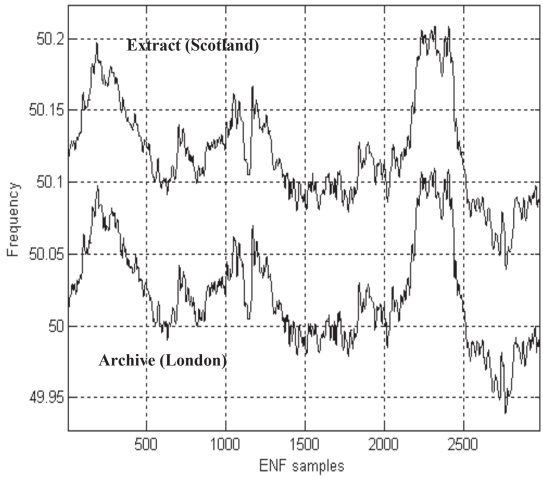

Figure 13: Automated match found using the techniques described. The extracted waveform has been offset by 0.1 Hz to aid visual comparison.

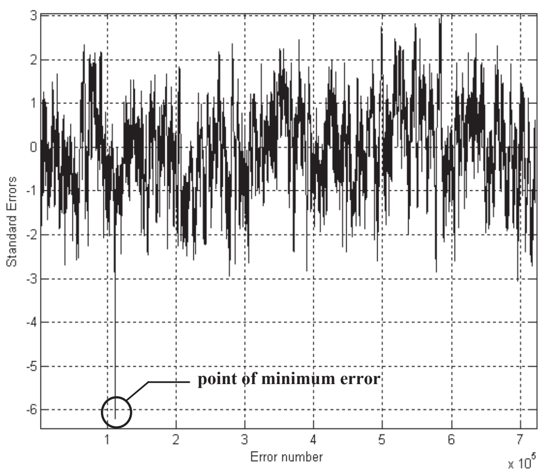

Figure 14: Standardised errors showing the point of the match being >6 standard errors below the mean.

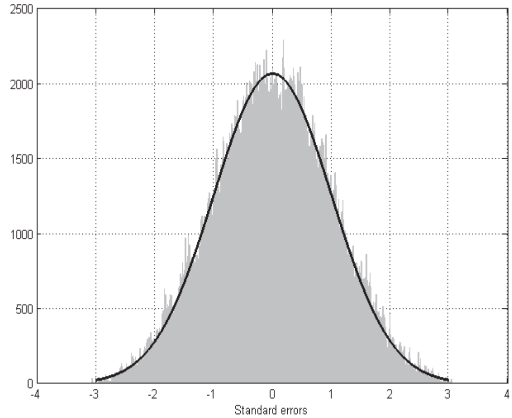

Figure 15: Histogram of errors with Gaussian overlay.

As the recording length diminishes the ENF pattern becomes less complex, the differentiation between lower error values therefore decreases and the probability of a match occurring by chance increases. However, if the extracted ENF signal is of good quality, reliable automated matches can still be obtained with relatively short recording lengths. As an example of this, a two-minute ENF extract has been matched to a 36 days archive file, with the results shown in Figure's 16, 17 and 18. Over this short recording, very good visual correlation can be seen between the archive and extracted ENF values (Figure 16). A high discrimination between the lowest error value and its nearest neighbours are shown in Figure 17. The error distribution is shown to be skewed (Figure 18) making statistical inference more difficult. However, transforming the results to Gaussian may be possible.

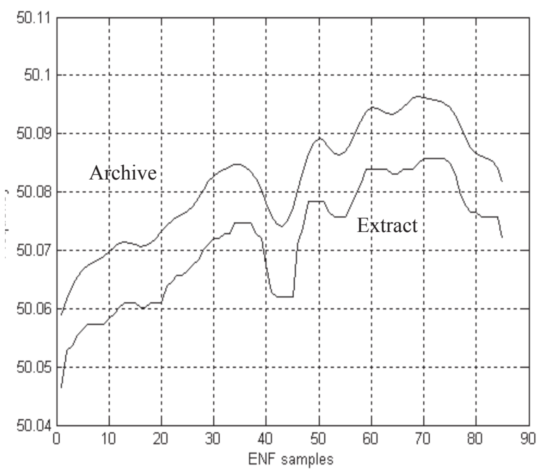

Figure 16: Automated match found using a 2 minute ENF extract. The archive waveform has been offset by 0.01 Hz to aid visual comparison.

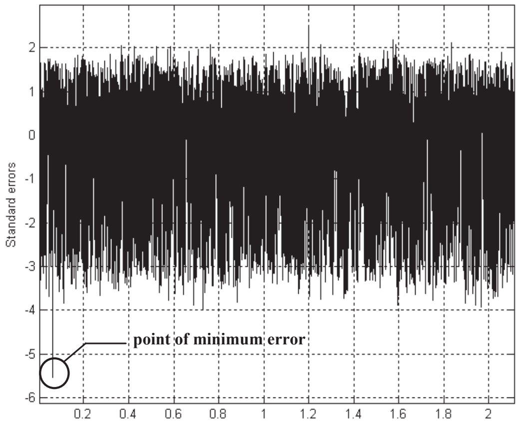

Figure 17: Even for a 2 minute section the minimum error value of the match is still well below its nearest neighbours.

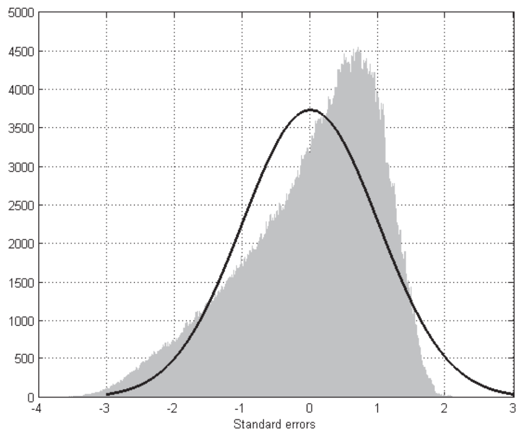

Figure 18: The overlaid Gaussian distribution highlights the skew in the error histogram.

Conclusions

This article further demonstrates the ENF criterion as a powerful methodology to help establish the integrity and authenticity of audio recordings. Validity of the criterion for use in mainland UK has been achieved by establishing ENF correlation over large geographical distances. Relatively simple signal processing procedures have been instigated allowing reliable automated date and time matching of extracted ENF data to an archived ENF file, even for a recording as short as two minutes in duration. An automated approach also has the important advantage of allowing statistical data to be derived indicating the strength of the match. Obviously, the reliability of a match diminishes with decreasing ENF-to-noise ratios and is the single biggest limitation of the ENF criterion. Although solutions have been proposed to help reduce the effects of noise, it is anticipated that future research will target the development of further robust and efficient ENF extraction algorithms.

About the author

Alan J. Cooper has 26 years experience with the Metropolitan Police Forensic Audio Laboratory where he is currently the technical manager. His duties are split between research/development and casework. He has a PhD in statistical signal processing relating to the detection of copied digital recordings for forensic purposes. Research interests primarily relate to technical issues concerning forensic audio authenticity examinations. He is a Chartered Scientist, Chartered Engineer and Chartered Physicist, is a member of the IAFPA, the Institute of Physics, the Institute of Acoustics, the Audio Engineering Society and is a Fellow of the Institution of Engineering and Technology.

Acknowledgements

The author would like to thank Dr Catalin Grigoras of the National Institute of Forensic Expertise in Bucharest Rumania for introducing him to the ENF criterion and for the useful discussions and exchange of ideas over the past several years.

The author is also grateful to the Parliamentary Office of Science and Technology for kind permission to reproduce the UK high voltage electric network diagram (Figure 1) from 'Electricity in the UK *Postnote* 280'.

Thanks also for the helpful comments of Robin How and James Gott and to the two anonymous referees.

References

Abe, M. and Smith III, J. O. (2004) Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of FFT magnitude peaks. 117th Audio Engineering Society Convention, San Francisco, CA, USA. Convention paper 6256.

Allen, J. B. (1978) Short term spectral analysis, synthesis, and modification by discrete fourier transform. *IEEE Transactions on Acoustics Speech and Signal Processing* 25(3): 235–238.

Boss, D., Gfroerer, S. and Neoustroev, N. (2003) A new tool for the visualisation of magnetic features on audiotapes. *The International Journal of Speech Language and the Law* 10(2): 255–276.

Bouten, J., Van Rijsbergen, M. and Donkers, S. (2007) Derivation of a transfer function for imaging polarimetry used in magneto-optical investigations of audio tapes in authenticity investigations. *Journal of the Audio Engineering Society* 55(4): 257–265.

Brixen, E. B. (2007) Techniques for the authentication of digital audio recordings. 122nd Audio Engineering Society Convention, Vienna, Austria.

Brixen, E. B. (2008) ENF: quantification of the magnetic field. 33rd Audio Engineering Society International Conference, *Audio forensics – theory and practice*, Denver, Colorado, USA.

Cooper, A. J. (2005) The significance of the Serial Copying Management System (SCMS) in the forensic analysis of digital audio recordings. *The International Journal of Speech Language and the Law* 12(1): 49–62.

Cooper, A. J. (2008a) Detection of copies of digital audio recordings produced using analogue interfacing. *The International Journal of Speech Language and the Law* 15(1): 67–95.

Cooper, A. J. (2008b) The Electric Network Frequency (ENF) as an aid to authenticating digital audio recordings: an automated approach. 33rd Audio Engineering Society International Conference, *Audio forensics – theory and practice*, Denver, Colorado, USA.

Czyzewski, A., Ciarkowski, A., Kaczmarek, A., Kotus, J., Kulesza, M. and Maziewski, P. (2007) DSP techniques for determining 'wow' distortion. *Journal of the Audio Engineering Society* 55(4): 266–284.

Dean, D. J. (1991) *The Relevance of Replay Transients in the Forensic Examination of Analogue Tape Recordings*, *Publication No 16*. London: Police Scientific Development Branch, Home Office.

Desainte-Catherine, M. and Marchand, S. (2000) High precision fourier analysis of sounds using signal derivatives. *Journal of the Audio Engineering Society* 48(7/8): 654–667.

French, J. P. (1993) Authenticity examinations of audio recordings. *Proceedings of the Institute of Acoustics, Reproduced Sound* 9 15(7). Liverpool: Rayross.

Grigoras, C. (2005) Digital audio recording analysis: the Electric Network Frequency (ENF) criterion. *The International Journal of Speech Language and the Law* 12(1): 63–76.

Grigoras, C. (2007a) Application of ENF criterion in forensic audio, video, computer and telecommunication analysis. *Forensic Science International* 167: 136–143.

Grigoras, C. (2007b) Application of ENF method in authentication of digital audio and video recordings. 123rd Audio Engineering Society Convention, New York, USA. Convention paper 1273.

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

Harris, F. J. (1978) On the use of Windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE* 66(1): 51–83.

Huijbregtse, M. and Geradts, Z. (2009) Using the ENF criterion for detecting the time of recording of short digital audio recordings. Proceedings: Lecture Notes in Computer Science 5718, Springer: 116–124.

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* 155: 165–171.

Koenig, B. E. (1990) Authentication of forensic audio recording. *Journal of the Audio Engineering Society* 38(1/2): 3–33.

National Grid Company. Electricity real time operational data. http://www.nationalgrid.com/uk/electricity/data/Realtime/ Accessed 24 December 2008.

Papoulis, A. and Pillai, S. U. (2002) *Probability, Random Variables and Stochastic Processes*. (Fourth Edition.) New York: McGraw-Hill.

Parlimentary Office of Science and Technology (2001) UK Electricity Networks *Postnote* 163. http://www.parliament.uk/post/pn163.pdf Accessed 24 December 2008.

Parlimentary Office of Science and Technology (2007) Electricity in the UK *Postnote* 280. http://www.parliament.uk/documents/upload/postpn280.pdf Accessed 24 December 2008.

Rife, D. C. and Boorstyn, R. R. (1974) Single tone parameter estimation from discrete-time observations. *IEEE Transactions on Information Theory* 20(5): 591–598.

Sanders, R. W. (2008) Digital audio authenticity using the electric network frequency. 33rd Audio Engineering Society International Conference, *Audio forensics – theory and practice*, Denver, Colorado, USA.

Sidhu, T. S. (1999) Accurate measurement of power system frequency using a digital signal processing technique. *IEEE Transactions on Instrumentation and Measurement* 48(1): 75–81.

Smith, J. O. and Serra, X. (1987) PARSHL: an analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation. *Proceedings of the International Computer Music Conference*, Tokyo.