Using the ENF criterion for determining the time of recording of short digital audio recordings

Maarten Huijbregtse, Zeno Geradts

[1] Netherlands Forensic Institute, Departement Digital Evidence and Biometrics, Laan van Ypenburg 6, 2497 GB DEN HAAG Netherlands
Z.geradts@nfi.minjus.nl

Abstract. The Electric Network Frequency (ENF) Criterion is a recently developed forensic technique for determining the time of recording of digital audio recordings, by matching the ENF pattern from a questioned recording with an ENF pattern database. In this paper we discuss its inherent limitations in the case of short – i.e., less than 10 minutes in duration – digital audio recordings. We also present a matching procedure based on the correlation coefficient, as a more robust alternative to squared error matching.

Keywords: ENF, authentication, integrity

1 Introduction

Electric networks operate at their own specific frequency: the Electric Network Frequency (ENF)[1]. However, due to unbalances in production and consumption of electrical energy, the ENF is known to fluctuate slightly over time rather than being stuck to its exact set point (figure 1). The fluctuation pattern is the same throughout the entire network [2] [3].

Digital recording equipment – both mains and battery powered – can pick up the ENF[2], which ends up as an extra frequency component in the recorded audio file [2] [3] [4]. By band pass filtering the audio signal, the ENF can be isolated and its pattern can be retrieved. Under the assumption that the ENF fluctuations are random, this effectively puts a time-stamp on the audio recording: the ENF pattern is unique for the time at which the recording was made.

The ENF criterion

One of the challenges in authenticating digital audio evidence is to gain insight into its time of recording [5]. A technique known as the *ENF criterion* [2] uses the aforementioned ENF fluctuation to achieve this[3]. By comparing the recorded ENF pattern to a database ENF pattern from the same electric network, it is possible to:

1) verify (or falsify) a questioned time of recording, or
2) determine an unknown time of recording.

A visual comparison of the recorded and database ENF patterns is often adequate for the first case, while an (automated) search routine is necessary for the latter, to locate the best match between recorded and database pattern[4].

1 The main part of continental Europe is served by one large electric network, controlled by the UCTE [1]. Its ENF is set at 50 Hz.

2 Claims are that recording equipment's microphones are sensitive to the power socket signal (when mains powered) and the electromagnetic fields emanating from nearby power lines (when battery powered). A thorough investigation of recording equipment for which these claims hold is, however, lacking.

3 See [2] for other applications of the ENF criterion.

4 In this paper, we focus on using the ENF criterion for determining an unknown time of recording.

Paper outline

Up until now, the ENF criterion has mostly been used with long recordings – i.e., approximately one hour in duration [2] [3] [4]. However, the amount of digital audio evidence (often accompanied by video footage) of short duration – i.e., ten minutes or less – is increasingly prominent with the advent of audio and video recording capabilities in consumer products (e.g., cell phones and digital cameras).

In the second half of this paper (sections 4 and 5) we will discuss the limitations of using the ENF criterion with short recordings, and show examples of erroneous determination of the time of recording when using a minimum squared error-based matching procedure. We present a maximum correlation coefficient-based matching procedure as a more robust alternative.

In the first half (sections 2 and 3), we will describe our means of building an ENF pattern database and extracting the ENF pattern from a digital audio recording.

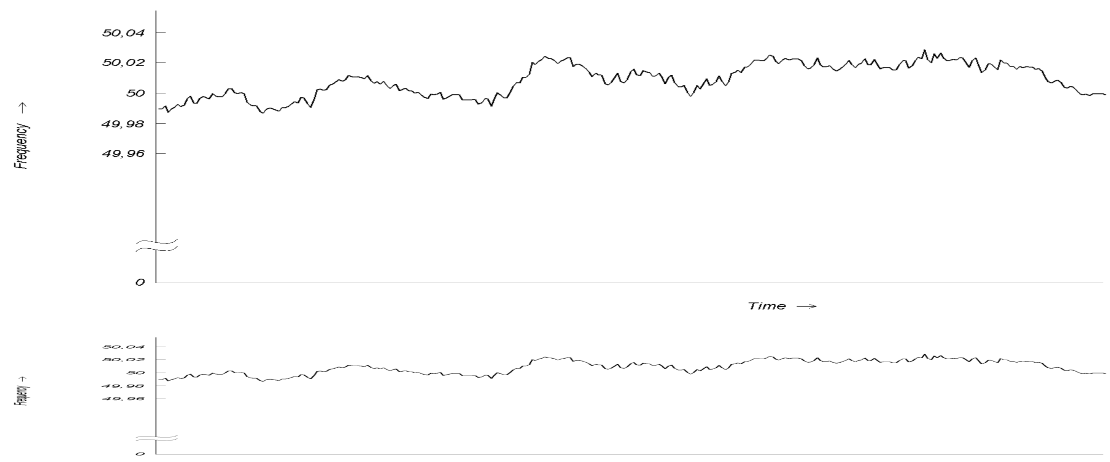


Fig 1 : ENF fluctuation over time.

2 ENF pattern database

Since the ENF is the frequency at which voltage levels in an electric network oscillate, it is possible to obtain the ENF pattern by analyzing the voltage level signal, e.g., from a power socket. In our setup, we fed this signal – in attenuated form – to a PC sound card that was set to a sampling frequency of 8000 Hz. The sampled signal $x[n]$, with the index *n* starting at 1, can be modeled as:

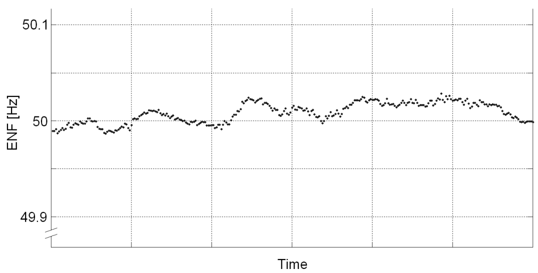
$$x[n] = k \cdot V\left(\left(n-1\right)T_s + t_1\right) \qquad \textbf{(1)}$$

where $V(t)$ denotes the voltage level at time t, k is a factor representing the attenuation, T_s the sampling period (= $1.25 \cdot 10^{-4}$ s) and t_1 the time of the first sample.

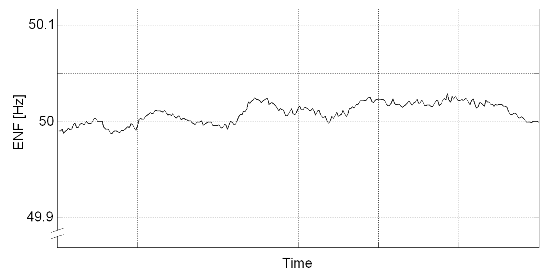
We used the method of zero crossings, mentioned by Grigoras [2], for analysis of $x[n]$. The idea is to treat the signal as sinusoidal, although this is not strictly true since its frequency – the ENF – varies slightly over time. For a sinusoidal signal, the time τ between two consecutive zero crossings equals half the oscillation period, so that its inverse equals twice the oscillation frequency f :

$$f = \frac{1}{2\tau} \qquad \textbf{(2)}$$

We determined the times of zero crossings by linear interpolation between samples $x[k]$ and $x[k+1]$ that differ in sign, and calculated the difference between two consecutive times of zero crossings to obtain values for τ . The corresponding values for f were calculated using (2) and averaged for every second of signal. This finally results in a series of frequency values – i.e., the ENF pattern – with a time resolution of 1 second (figure 2a). For visual clarity, ENF patterns are often depicted as continuous (figure 2b).



a)



b)

Fig.2 a) ENF pattern as a series of ENF values b) Continuous ENF pattern, obtained by interpolating ENF values

3. ENF pattern extraction from digital audio recording

We have adopted the method presented by Cooper [4] for extracting the ENF pattern from a digital audio recording. We shall cover this method briefly here, since Cooper's paper offers an excellent and comprehensive description.

The basic steps are:

- *Signal decimation* – Many digital audio recordings are recorded at high sampling frequencies – e.g., 44100 Hz. To detect the ENF, which is approximately 50 Hz, much lower sampling frequencies are allowed. The audio file is thus decimated to a sampling frequency of 300 Hz, which significantly reduces computational time.
- *Band pass filtering* – The frequencies of interest are around 50 Hz, so the decimated audio file is digitally band pass filtered from 49.5 Hz to 50.5 Hz to isolate the ENF.
- *Short Time Fourier Transform (STFT)* – In discrete time STFT analysis, a signal is divided into J partly overlapping frames (figure 3) for which, after windowing and zero-padding, the frequency spectrum is calculated via a Discrete Fourier Transform (DFT). The jump H (in samples) between frames determines the time resolution of the final ENF pattern, while the amount of overlap $M - H$ affects its smoothness. In our specific case, we have chosen $H = 300$ so that the extracted ENF pattern time resolution equals that of the database – i.e., 1 second. Each frame was windowed with a rectangular window and zero-padded by a factor of 4.
- *Peak frequency estimation* – For each frequency spectrum[5], the frequency with maximum amplitude is estimated. As it is unlikely that this 'peak frequency' coincides exactly with a DFT

5 Actually, we used the *log power spectrum*, defined as $\log_{10}\left|X[f]\right|^2$, where $X[f]$ is the frequency spectrum.

frequency bin, quadratic interpolation around the bin with maximum amplitude is performed. The estimated peak frequency is stored as the ENF value for the corresponding frame, so that we end up with an extracted ENF pattern of *J* ENF values.

4 Matching by minimum squared error

Calculating the squared difference ('error') between two vectors is a common approach in determining their equivalence: the smaller the squared error, the more both vectors are alike. The squared error E for two length L vectors x and y is defined as:

$$E = \sum_{i=1}^{L} \left(x[i] - y[i] \right)^2 \qquad \textbf{[3]}$$

When determining the time of recording using the ENF criterion, we have in general one longer vector (the database ENF pattern d) and one shorter vector (the recorded ENF pattern r). The approach is then to calculate a vector e of squared error values, according to:

$$e[k] = \sum_{i=1}^{R} \left(r[i] - d[i+k-1] \right)^2 \qquad \textbf{(4)}$$

in which R is the length of the recorded ENF pattern, while the index k runs from 1 to $D - R + 1$; D being the length of the database ENF pattern. The minimum value in e determines the location of the best match between recorded and database pattern, and hence the time of recording.

Ideally, the recorded ENF pattern and its corresponding database pattern are exactly equal and the minimum error value would be zero. In practice, however, this is not the case. The reliability of the ENF criterion in determining the right time of recording is therefore limited by the occurrence of similar patterns *within the database itself* – i.e., ENF patterns with squared errors in the same range as 'typical' squared errors between recorded and corresponding database ENF pattern.

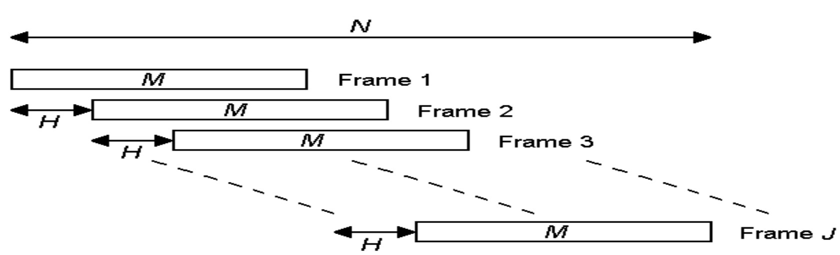


Fig. 3 Division of a signal of length N into J partly overlapping frames

Database analysis

In an experiment, we took roughly 1.5 years of ENF data[6] and calculated the (root mean) squared error[7] between two randomly picked, non-overlapping pieces of 600 ENF values ($\hat{=}$ 10 minutes). By repeating this one million times, we were able to picture the approximate distribution of root mean squared (rms) errors between ENF patterns of length 600 within the database (figure 4). Similar experiments were run for

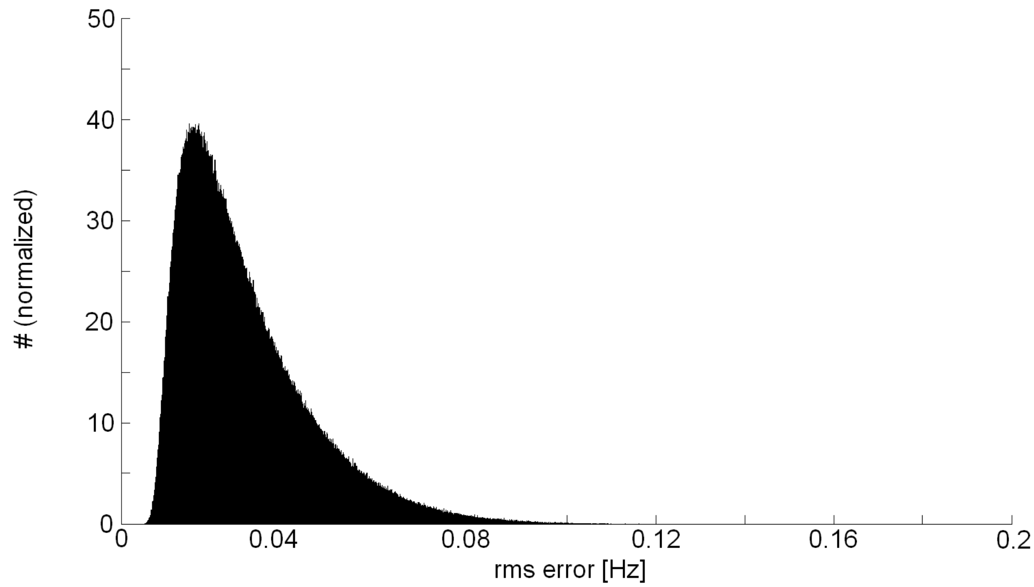


Fig 9 : Normalized histogram of squared errors between database pieces, 600 ENF values in length

patterns of 60, 120, 240 and 420 ENF values.
The most interesting part of the histogram in figure 4 is near zero: the smallest (observed) rms error within the database ENF pattern. For length 600, we found this smallest rms error to be about 0.0040 Hz.

We thus conclude that the minimum rms error between a recorded ENF pattern of length 600 and a (large) database should be 'well below' 0.0040 Hz for a reliable determination of the time of recording.

Table 1 lists the observed smallest rms errors for all experiments. As can be expected, the error increases for longer ENF patterns: the longer a pattern, the less likely it will have a similar counterpart over its whole length within the database[8].

---

6 Collected as described in section 2, at the Netherlands Forensic Institute (The Hague, The Netherlands) from September 2005 until February 2007. ENF values were stored minute-by-minute in plain text files (i.e., 60 values per file).

7 Following the notation of equation (3), the root mean squared error E_{rms} is defined as $E_{rms} = \sqrt{\frac{\sum_{i=1}^{L} \left(x[i] - y[i]\right)^2}{L}}$. Conclusions are independent of a choice for E or E_{rms} as the dissimilarity measure.

8 It is therefore that the ENF criterion works well for audio recordings of long duration (as confirmed by some of our experiments not mentioned here). In this case, the rms error between recorded and corresponding database ENF pattern is almost certainly much smaller than those found within the database itself.

Table 1 : Smallest observed rms errors within 1.5 years of ENF database

ENF pattern length	Smallest observed rms error [Hz]
60	0.0007
120	0.0015
240	0.0020
420	0.0035
600	0.0040

Test recordings

For a second experiment, we took an "American Audio Pocket Record" portable digital audio recorder and set it up to be mains powered. We made a total of 70 recordings with durations of 60, 120, 240, 420 and 600 seconds (i.e., 14 recordings for each duration). The exact times of recording were known and the audio files – sampled at 44,1 kHz – were stored in lossless WAV format.

The ENF pattern from each recording, extracted as described in section 3, was compared to a small database consisting of two weeks of ENF data, including the period of recording. Results are summarized in table 2.

Table 2: Test recording results for minimum root-mean-squared error matching

Recording duration	Correct time estimate for…	Minimum rms error ranging from…
60 s	0 out of 14 recordings	0.0008 Hz to 0.0028 Hz
120 s	0 out of 14 recordings	0.0018 Hz to 0.0037 Hz
240 s	2 out of 14 recordings	0.0033 Hz to 0.0049 Hz
420 s	10 out of 14 recordings	0.0045 Hz to 0.0055 Hz
600 s	14 out of 14 recordings	0.0045 Hz to 0.0055 Hz

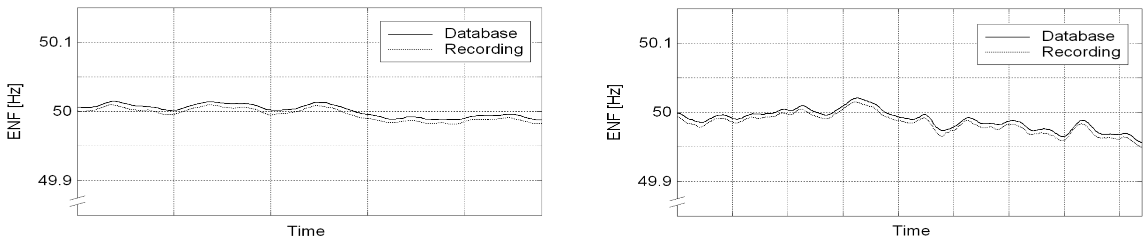
It is seen that the ENF criterion failed in correctly estimating the time of recording for 44 out of the 70 recordings. Moreover, the found minimum rms errors are all above the values mentioned in table 1. Comparison to a larger database could thus have resulted in even less satisfying results.

Matching by maximum correlation coefficient

Figure 5 shows a main reason for the failure of the ENF criterion: our recorded ENF patterns have a slight offset compared to the database pattern – a phenomenon also noted by Kajstura et al [3]. In general, this cannot be known beforehand and thus the matching procedure should be robust to this type of behavior.

We propose matching based on equivalence of shape, by using the correlation coefficient. Following the notation of equation (3), the correlation coefficient ρ between two vectors is defined as:

$$\rho = \frac{\sum_{n=1}^{L} (x[n] - \overline{x})(y[n] - \overline{y})}{(L-1)\sigma_x \sigma_y} \qquad \textbf{(5)}$$



a) b)

Fig 5 : Recorded ENF patterns lie consistently below the corresponding database patterna) Example of a recording 240 s in duration b) Example of a recording 420 s in duration

where the horizontal bars and sigmas denote the averages and the standard deviations of the vectors respectively. ρ can run from -1 to +1; the closer the value is to +1, the more both vectors are alike in shape. When comparing a recorded ENF pattern to a database, we thus search for the *maximum correlation coefficient* between recorded and database pattern.

Database analysis

As with minimum squared error matching, the reliability of a maximum correlation coefficient-based matching procedure will be limited by high correlations within the database itself. We have repeated the first experiment described in the preceding section, this time calculating correlation coefficients instead of rms errors. Here, we are interested in the *largest* observed values, which are listed in table 3[9].

Table 3: Largest observed correlation coefficients within 1.5 years of ENF database

ENF pattern length	**Largest observed correlation coefficient**
60	0. 9980
120	0.99
240	0.9870
420	0.9820
600	0.9850

Test recordings

Matching the same 70 test recordings with the same database by a maximum correlation coefficient search, yielded the results mentioned in table 4. Correct time estimation is significantly improved, with only 3 failures out of 70. Also, from duration of 240 seconds onwards, the maximum correlation coefficients all lie above the values mentioned in table 3. This suggests that even comparisons to a larger database would have resulted in correct time estimates.

Table 4: Test recording results for maximum correlation coefficient matching

Recording duration	**Correct time estimate for…**	**Maximum corr. coeff. ranging from…**
60 s	12 out of 14 recordings	0.9758 to 0.9989
120 s	13 out of 14 recordings	0.9572 to 0.9980
240 s	14 out of 14 recordings	0.9893 to 0.9989
420 s	14 out of 14 recordings	0.9910 to 0.9992
600 s	14 out of 14 recordings	0.9945 to 0.9993

9 The higher value for length 600 compared to length 420 is probably due to using 'only' one million random pairs of database pattern: the experiment for length 600 just happened to come across a better matching pattern than the one for length 420.

6. Conclusion

We have shown that the reliability of the ENF criterion is inherently limited by similarities within the ENF pattern database to which the recording is compared. The possible presence of a frequency offset further increases the danger of erroneous determination of the time of recording – especially for recordings shorter than 10 minutes in duration in combination with a minimum squared error-based matching procedure. We have shown improvements by using a maximum correlation coefficient-based matching procedure.

References

1. Union for the Co-ordination of Transmission of Electricity, Homepage, http://www.ucte.org
2. Grigoras, C. : Digital audio recording analysis – the electric network frequency criterion, International Journal of Speech Language and the Law, vol. 12, no. 1, pp. 63-76 (2005)
3. Kajstura, M, Trawinska A, Hebenstreit, J: Application of the Electrical Network Frequency (ENF) Criterion – A case of a Digital Recording, Forensic Science International, vol. 155, pp. 165-171 (2005)
4. Cooper, A.J: The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings – an automated approach, Conference paper, AES 33rd International Conference, USA (2008)
5. Brixen, E.B: Techniques for the authentication of digital audio recordings, Presented at the AES 122nd Convention, Vienna (2007)