Exploring Power Signatures for Location Forensics of Media Recordings

IEEE Signal Processing Cup 2016

Andreas Triantafyllopoulos, Anastasios Foliadis, George Roustas, Ioannis Krillis, Fanouria Athanasiou, Maria Papaioannou

and.triantafy@gmail.com, anfoliadis@gmail.com, just0m3@gmail.com, eejohnkr@gmail.com, fanouria.ath@gmail.com,mar93pap@hotmail.gr
Electrical and Computer Engineering Department
University of Patras
Patras, Greece

Abstract-The estimation of location based on the time varying Electric Network Frequency (ENF) is a new emerging technology in Information Forensics. This requires the extraction of the ENF signal from multimedia recordings and a comparison with already known power grids from different locations. The decision making is possible using machine learning algorithms. In this report, we focus on ENF signal extraction and statistical modelling of ENF signals. We extract features and develop a classification system to accurately identify region-of-recording.

Table of Contents

1. Introduction

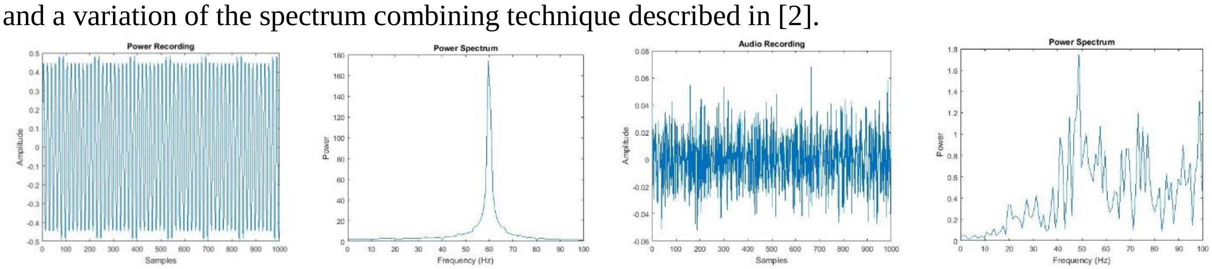
The ENF (Electrical Network Frequency) analysis is used as a new forensic science technique in order to estimate the location of an audio or video recording. The nominal value of ENF is either 50 or 60 Hz, depending on the country, with 60 used mostly in North America and Japan and 50 in most other countries. However, the ENF is not steady but fluctuates over time due to variations in the demand and supply of electric power. These variations present a generally consistent trend within the same grid. ENF signals consist of changes in ENF over time and can be extracted from recordings either directly from a power socket or using a portable audio recorder. These recordings are segmented and processed in order to produce an estimation of the frequency at each segment. It has been shown that signals extracted from video or audio recordings are similar to concurrently recorded clean power signals.

Applications using ENF signals include detection of tampering or modifications in a multimedia signal, time and location-of-recording validation and region-of-recording identification. Our work focuses on the later application. As mentioned, multimedia signals show similarities with power references from the same grid. Working under the assumption that variations of the ENF are consistent within the same grid and differ among distinct countries, we extract statistical features from the signals in our database and develop a classification system to accurately identify the region-of-recording.

Section 2 will provide an overview of the approaches used to extract the ENF signals from our recordings. Section 3 will provide details of the classification method used and our results. Finally, Section 4 will describe our implementation of the sensing circuit which is used for ENF acquisition.

2. Extraction of ENF signals

We worked on two different sets of recordings, some consisting of signals acquired through hardware connected directly to a power socket using a step down transformer, thus effectively measuring the power grid, and some acquired using a battery powered digital recorder. In the latter occasion the recorder is influenced by surrounding power sources and such recordings tend to be more noisy, requiring different signal extraction techniques. We proceed to describe two different methods for extracting ENF signals, a simple zero crossings algorithm that has been proven adequate for direct-from-power-socket recordings [6] and a variation of the spectrum combining technique described in [2].



(a) *(b)* *(c)* *(d)*

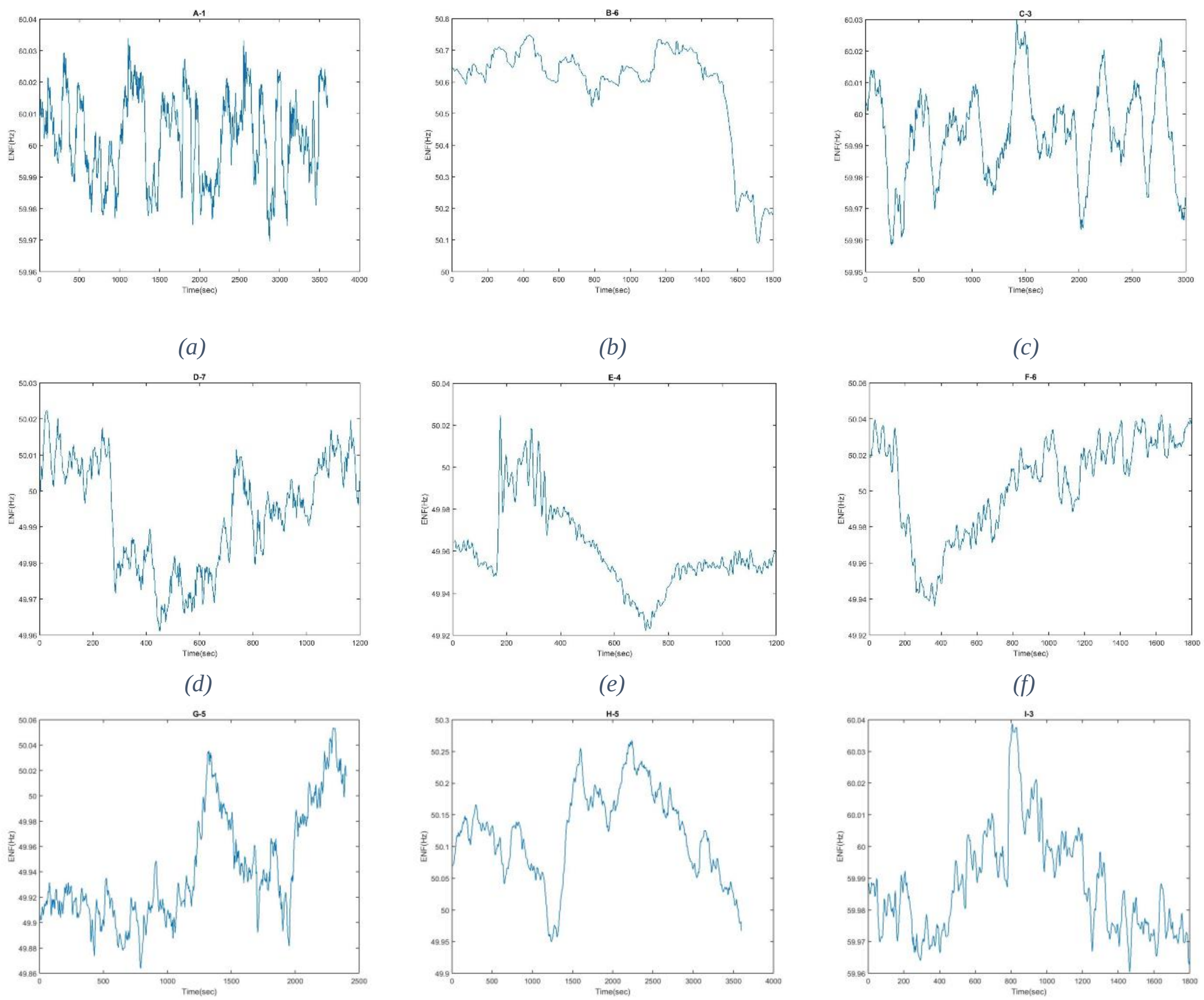
Figure 1 depicts one second of the recorded time signal and the corresponding spectrum for that second for a power (a, b) and an audio recording (c, d)

Our initial observation is that audio recordings tend to be more noisy than the ones made with a power recorder as is evident by both time and frequency domain plots, thus justifying our choice of different ENF extraction methods for different noise levels. We also observe that zero-noise recordings from the same grid present similar power levels around the ENF and its harmonics and that these levels vary among different grids. This variation is present in audio recordings as well, though it must be noted that power levels also differ between zero-noise and audio recordings of the same grid.

2.1 Zero crossings

Zero crossings is a method used by [6] that can produce a fast estimation of the frequency of a sinusoidal signal. Due to the time-varying nature of the ENF we apply our method to discrete overlapping frames, during which we consider the ENF to be stable. We empirically choose a time frame of 4096 samples, 1024 of which are overlapping. Our implementation consists of the following steps:

1. Estimate the central frequency of the signal, which is the frequency with the highest coefficient when the whole signal is decomposed in Fourier series.
2. Filter the signal with a band-pass Butterworth filter with a 10 Hz pass-zone centered on its central frequency to remove harmonic content not related to the ENF.
3. Divide the signal to overlapping frames of 4096 samples each, with overlap size 1024 samples.
4. For each frame we find all samples x[k], x[k+1] that differ in sign and linearly interpolate between them to find the time of crossing.
5. For two consecutive crossings we compute the difference between the respective times of crossing and acquire an estimation of the signal's period and corresponding frequency.
6. Average the computed frequencies to get an estimation of the ENF for that frame.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 3 Sample ENF signals extracted from the provided power recordings.

Figure 3 presents ENF signals extracted from the power recordings of our dataset using the afore-mentioned method. We can divide the grids of our database in two categories, based on their mean frequency; we have three grids with a mean frequency of 60 Hz and six with 50 Hz. Among the three grids with a base frequency of 60 Hz, grid A exhibits the lowest variations around its mean value.

2.2 Spectrum combining

In section 2 we showed that audio recordings are more noisy and have a different harmonic content than recordings directly from a power socket. As a result the zero crossings method discussed in section 2.a fails to correctly extract the ENF. We use a variation of the method used in [2] and produce an ENF estimation by combining base and harmonic spectral bands each with a corresponding weight. Our implementation differs from [2] in that we choose to use the chirp Z-transform to approximate the maximum coefficient and the corresponding. The chirp Z-transform is the Z-transform of a signal along an arbitrary spinal contour. The contour used in each step of our algorithm is a segment of the unit circle, whose limits are specified by a range of 1 Hz around the frequency estimation computed in the previous step. This allows us to calculate the coefficients inside the area of interest with increased accuracy.
Our implementation consists of the following steps:

1. Calculation of the basic frequency of the recording.
2. For the first eight harmonics of our signal, estimate the power density within a band close to the value of each harmonic. The power density outside of this narrow band within a range of 1 Hz is assumed to be noise. The boundaries of these bands for each harmonic increase accordingly. The resulting SNR of is used as combining weight. Only the two harmonics with the biggest weights are taken into account for the next steps to decrease computation time.
3. Segment the signal into frames of 4096 samples, 1024 of them overlapping.
4. Compute the chirp Z-transform for each segment around the two harmonics with the largest weights and use the frequency corresponding to the maximum value as the ENF estimation for that segment.
5. Apply median filtering on the resulting signal to remove outliers.

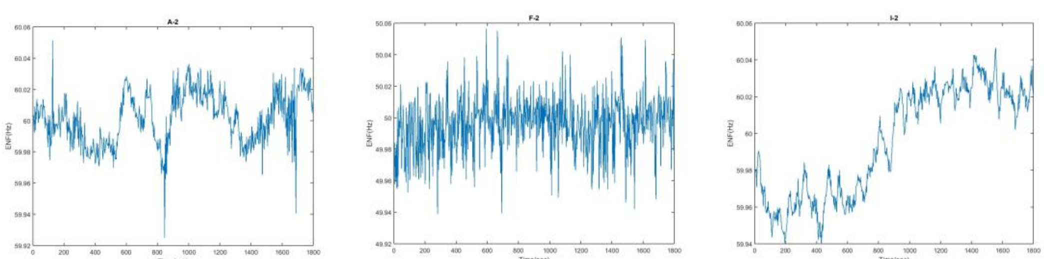
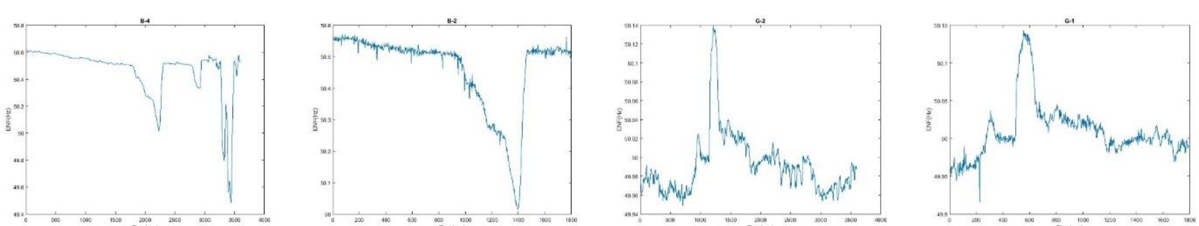


Figure 4 *Sample ENF signals extracted from audio recordings using the method described above.*



(a) *(b)* *(c)* *(d)*

Figure 5 *Sample ENF signals (b, d) extracted from audio recordings show definite similarities to clean power signals (a, c).*

3. Location classification/identification system

3.1 System Design

In this section we describe the classification system used to identify region-of-recording. Our dataset consists of 9 different grids, therefore we must develop a multiclass classification scheme that could accurately differentiate between them.

As a first step we must choose features that could quantitatively describe our grids. We use sets of 10 minute long segments of ENF signals. The features shown in Table 1 have been used in [1] and showed promising results. The mean value is chosen to separate classes of 50 and 60 Hz. Our initial observations on ENF signals suggest that the variance and range of each signal will be good predictors of its region-of-recording. We use a 9-level dyadic wavelet decomposition to get an approximation of the ENF signal and the details at different levels of resolutions. These features will help as capture the variations of the ENF among different grids.

An autoregressive (AR) model is used to statistically model our signal. [1] proposes an AR model of order 2. The equation that describes our signal in this context is:

$$s[n] = \alpha_1 s[n-1] + \alpha_2 s[n-2] + \upsilon[n]$$

We use α_1, α_2 and the variance of the innovation signal $\upsilon[n]$ as three feature values to distinguish ENF signals in how well they fit an AR model and in what manner.

Table 1 Features used in our classification model

Index	Feautures
1	Mean of ENF segment.
2	log (variance) of ENF segment.
3	log(range) of ENF segment.
4	log (variance) of approximation after 9-level wavelet analysis.
5-13	log(variance) of nine levels of detail signal computed through 9-level wavelet analysis
14-15	AR(2) model parameters α_1 and α_2
16	log(variance) of the innovation signal after AR(2) modeling.

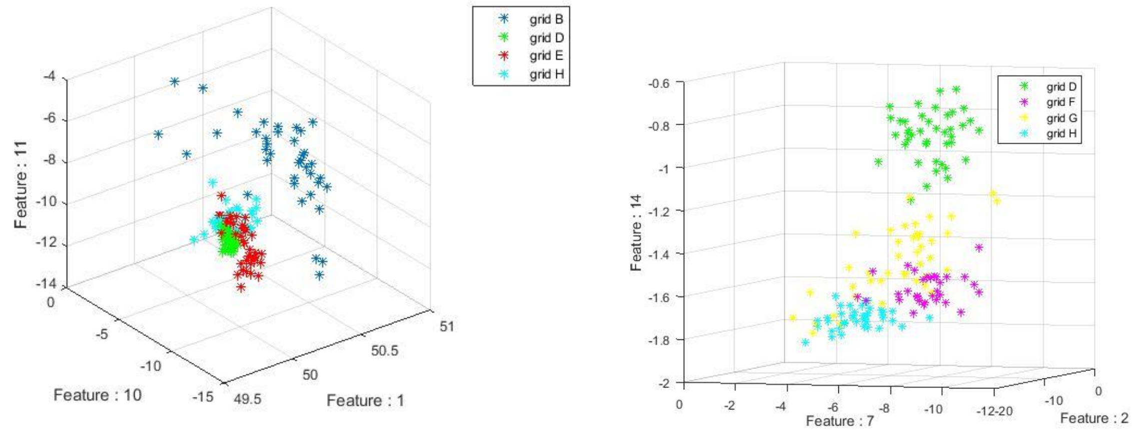
The 16 features extracted from our segments must be normalized in order to avoid any bias caused by difference in the order of magnitude of our features. We use a standard normalization scheme, subtracting the mean value and dividing by the standard deviation of each feature column:

$$\mu_k = \frac{1}{N}\sum_{i=1}^{N} f_i[k]$$

$$\sigma_k^2 = \frac{1}{N-1}\sum_{i=1}^{N}(f_i[k] - \mu_k)^2$$

$$f_i'[k] = \frac{f_i[k] - \mu_k}{\sigma_k}$$

Where we assume N to be the number of examples. The process is repeated for k = 1, 2,…, 16. The normalization parameters are stored to be applied to the testing examples.



(a) *(b)*

Figure 6 Feature values for instances of training data from 50Hz grids (a) Features 1, 10 and 11. (b) Features 2, 7 and 14

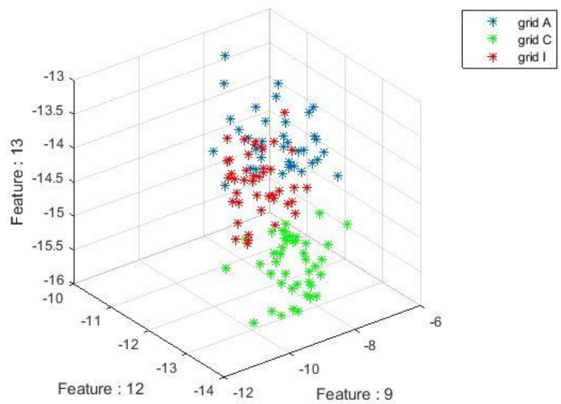


Figure 7 Feature values for instances of training data from 60Hz grids (Features 9, 12 and 13).

In figures 6, 7 we can see examples of features taken from different classes. We can see that 60 Hz grids form generally separable but overlapping clusters. Some of the 50 Hz grids also form distinct clusters, with grid B being easily distinguishable from the rest. Grids D, F, G, H can also be linearly separated as seen in figure 6b. However, since there is some amount of overlap it would be better to use a non-linear classifier that would better differentiate between different classes.

As mentioned, our database consists of both clean and noisy ENF signals. Previous work in this area [1] has shown that classification works better when a common classifier is used, trained on power and multimedia signals. We were able to confirm this fact, with common classifiers performing consistently better, yielding a lower error in cross-validation and when tested with the practice dataset. What is more, we were provided with very few multimedia recordings to build a separate classifier. Henceforth, our training dataset consists of features extracted from clean and noisy ENFs unless otherwise mentioned.

Our supervised learning classifier makes use of Support Vector Machines (SVM). We use one-vs-one classification and train a binary classifier for every possible pair of classes. In our implementation we use MATLAB's Statistics and Machine Learning toolbox to train our classifier and the Optimization toolbox to optimize its parameters. We train a multiclass SVM with the fitcecoc function and then pass it as argument to fminsearch, a function that performs unconstrained nonlinear optimization on a multi-variate scalar function, in order to minimize the cross-validation error. When initialized with a polynomial kernel of 2nd degree, which we assumed would adequately handle the nonlinearities present in our feature space, and default values specified for the fitcecoc function, the optimization algorithm converges to a polynomial kernel of degree 2.0005 indicating that our hypothesis was correct. All other parameters are optimized as well. The minimum 10-fold cross-validation error for our training dataset was 86.12%. When tested on the practice dataset we achieved a 90% classification rate. MATLAB also provides the ability to compute the posterior probabilities for our classifier, which serve as a confidence level for the classifier's decision.

Due to the fact that our database contains varying amounts of recordings for each grids, our system would be more biased towards grids with more training examples. In order to counter this fact we used a weighted SVM. SVMs include a cost parameter, which is the penalty for misclassifying an example. This cost is the same for all classes by default, but in weighted SVM it depends on the weight of each class. We choose larger cost values for smaller classes by manipulating the weights as follows:

$$w_j = \frac{N_{min}}{N_j}, \textit{ for } j = 1, 2, \ldots, M \textit{ and } Nmin = minNj$$

3.2 Classification results

Our method for classifying the practice and test dataset recordings consists of two steps. First, we must correctly extract the ENF from the given recordings. Since we are using different methods to estimate the ENF depending on whether we have a clean power signal or a multimedia recording, it is essential to differentiate between the two. A simple but effective method is to calculate the signal power over a band of frequencies that do not contain the frequencies 50 and 60 Hz and their harmonics. If the computed power is lower than an empirical threshold then the recording is assumed to be a clean ENF signal. This method has worked correctly for all the recordings in the three databases we worked with. After the recording's noise conditions are estimated we extract the ENF signal using the appropriate algorithm as mentioned in section 2. Consequently, we segment the signal in 10-minute frames and extract the features discussed in section 3. After appropriate normalization, we pass the testing example through our multiclass SVM and acquire a class prediction and posterior probabilities for all classes, which act as a confidence level for our classification. If the probability for the predicted class is lower than 0.45 then we classify the corresponding recording as not belonging to any of the 9 grids used for training.

Table 2 Confusion Matrix of our classification (percentages)

Grids	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
A	83.3333	0	0	0	0	0	0	0	16.6667
B	0	100	0	0	0	0	0	0	0
C	4.1667	0	93.75	0	0	0	0	0	2.0833
D	0	0	0	93.75	0	0	6.25	0	0
E	0	0	0	2.1277	82.9787	2.1277	12.7660	0	0
F	0	0	0	0	2.7027	83.7838	10.8108	2.7027	0
G	0	0	0	2.0833	2.0833	10.4167	70.8333	14.5833	0
H	0	0	0	0	0	2.1277	14.8936	82.9787	0
I	14.2857	0	2.408	0	0	0	0	0	83.6735

Table 2 shows the classification percentages for the training dataset using 10-fold cross-validation. It is evident that our classification systems accurately differentiates between grids of 50 and 60 Hz. This is a result of incorporating the mean ENF value as a feature. From the 60 Hz grids, C has the least overlap with the other two, as shown in figure 7, and therefore can be more easily identified by our classifier. We also achieve a 100% classification rate for grid B, which, as seen in figure 6, forms a distinct cluster from the other grids. What is more, none of the other grids are mistaken for grid B. In contrast, most misclassification errors are related with grid G, which is correctly classified only 70.83% of the time and is the grid most commonly mistaken for other grids.

We then tested our classifier on the practice dataset. Using this scheme, we obtain a 90% classification rate. We further analyzed our results testing the error rate for audio and power recordings separately. Our classifier achieves a 95% correct identification on audio recordings and an 86.6% rate on power recordings of the practice dataset. This was the best score we obtained for multimedia recordings after testing various classifications systems and led us to conclude on using an SVM trained on both power and audio recordings.

We then proceeded to pass the testing dataset through our classifier using the same method as the practice dataset. To sum up, for each recording we estimate the noise level, estimate the ENF signal using the appropriate algorithm, either zero crossings or spectrum combining, extract and normalize features and pass them through the trained classifier obtaining a prediction. If the confidence value is lower than 0.45 we assign the recording a None-of-the-above label, i.e. it does not belong to any of the 9 classes used for training. The results obtained for the training dataset are:

NNDCD,NNDAF,INGBF,BFCEH,DHHNG,FFEAI,DNFHI,IECBD,ENIBG,FGNAG,IINCG,HAEFC,CCFDG,CEIGI,EICEN,NEBHA,DINCG,AABIH,CNDBA,HBFBB

A more detailed analysis of these results entailing the confidence levels for each label is contained in the “testing_results.txt” file included in our submission.

4. Circuit design and data analysis for ENF acquisition

To record the power grid signal directly from a power socket we used the sensing circuit shown in figures 8, 9. The components used were a transformer, a glass fuse, a voltage divider, a high-pass filter and an anti-aliasing filter. The transformer decreases the voltage and provides galvanic isolation from the source to make further processing possible. For safety reasons we used a 400mA glass tube fuse. The voltage divider creates reference voltage of about 3Vp. The purpose of the high-pass filter is to offer high resistance to DC signals, so that only the AC component appears on the output with a cutoff frequency around 32Hz, while the anti-aliasing limits the bandwidth to 500 Hz. We used a potentiometer (R4) in order to be able to adjust the cutoff frequency of the anti-aliasing filter depending on the sampling rate of interest (R4-33k, 1 kHz). The signal is routed to the sound card through a 3.5 mm jack and can be recorded using appropriate software. In our implementation we used Audacity. We used a sampling rate of 1 kHz, consistent with the recordings of the provided dataset. The power consumption was calculated at 56 mWatt.

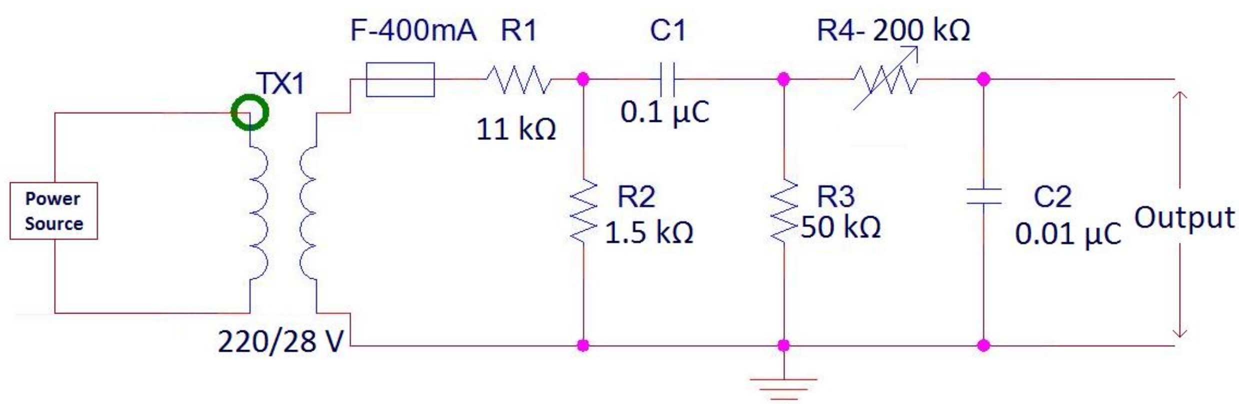


Figure 8 Schematic diagram of the sensing circuit

Figure 9 Photo of our sensing circuit

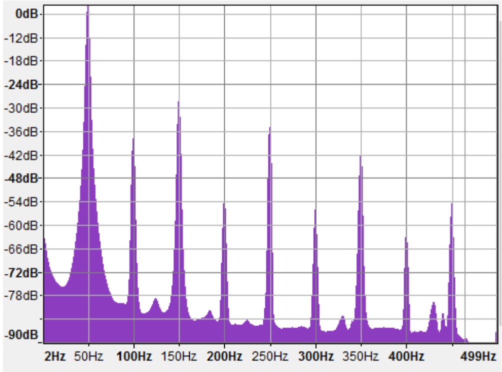


Figure 10 Frequency analysis of one of our recording

A frequency analysis of the data we recorded before their processing is shown in figure 10. Peaks at 50 Hz and higher harmonics indicate that the ENF signature is present in our recordings and the high SNR levels around those peaks suggest that the zero crossings algorithm described in section 2.a will yield appropriate results.

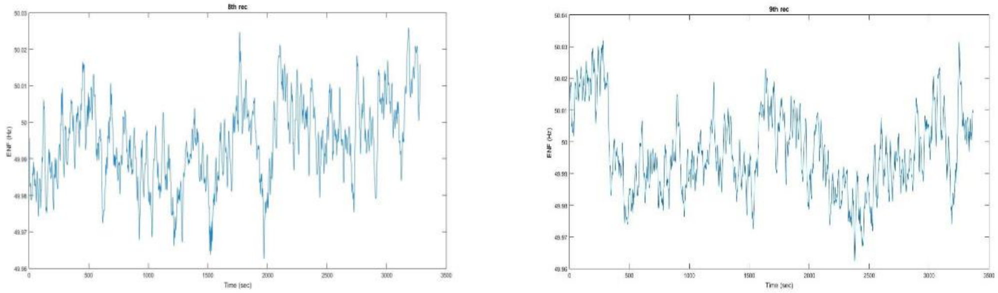


Figure 11 ENF examples of our grid

The ENF of our grid is quite similar to that of the most ENFs provided to us. The frequency is steady without fluctuations bigger than 0.1 Hz. Yet certain differences appear with those provides to us, whose frequency is more prone to changes.

We also used the power recordings of our country and tried to classify them using the system discussed in section 3. Detailed results are included in the "RecordingClassification.txt" file of our submission. We segmented the 10 hours of recordings to 10 minute frames and passed them through the classifier using the same approach as with the test and practice datasets. The Greek grid is mostly misclassified into belonging to grids G and F. Some segments are also confused with grids D and E, but with lower frequency and generally lower confidence. Only a small portion is classified as not belonging to any of the 9 grids used for training. This indicates that our grid shows sufficient overlap with grids F and G, which is problematic for the classification. The overlap with D and E is considered to be in acceptable levels and should be attributed to expected errors of any machine learning system.

5. Conclusion

In this report we present our region-of-recording classification results on the dataset provided to participants of the 2016 IEEE Signal Processing Cup Competition. We developed two methods to accurately extract ENF signals from recordings based on their respective SNR levels. We showed that these methods can procure ENF signals with a high degree of similarity within the same grid. This allowed us to develop a classification system based on a statistical modelling of these signals to correctly identify region-of-recording. The results obtained through cross-validation and on the provided practice dataset are satisfactory and in accord with results presented in [1].

As part of the competition we constructed an ENF sensing circuit presented in section 4. We were able to detect the ENF signature of the Greek grid using the same method chosen for clean recordings in the given dataset. This fact is an indication of our sensing circuit's performance. To further test and compare the Greek grid to the others in our databased we passed it through the classifier developed in section 3. However, the results were not appropriate, as our systems to identify the Greek grid as not belonging to the classes used for training. This raises concerns about our classification system's ability to respond to grids not belonging to the datasets we used and should be further investigated.

For our future work, we intend to continue working on ENF extraction and region-of-recording classification. We believe that there is potential for algorithm improvement in ENF extraction from noisy recordings, especially in recordings with low SNRs. We would also like to investigate the possibility of identifying the region-of-recording within the same grid using simultaneous recordings from different cities that belong to different sub-grids.

Acknowledgments

The authors would like to thank Dr. Athanasios Skodras for his invaluable guidance, assistance and support.

References

[1] Hajj-Ahmad, A.; Garg, R.; Min Wu, "ENF-Based Region-of-Recording Identification for Media Signals," in *Information Forensics and Security, IEEE Transactions on* , vol.10, no.6, pp.1125-1136, June 2015

[2] A. Hajj-Ahmad, R. Garg, and M. Wu, "Spectrum combining for ENF signal estimation," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 885–888, Sep. 2013.

[3] C. Grigoras, "Application of ENF Analysis in Forensics Authentication of digital Audio and Video Recordings" J.Audio Engineering Societ, vol. 5 , no. 9, pp. 643-661, Sept.2009.

[4] 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)

[5] ENF: Power Signature for Information Forensics. [Online resource]. Available: https://www.mast.umd.edu/index.php/enf-menu, accessed Jan. 2016

[6] Garg, R., Varna, A., & Wu, M. ,"Seeing ENF: natural time stamp for digital video via optical sensing and signal processing," *Proceedings of the 19th ACM International Conference on Multimedia*, 23–32 (2011)

Andreas Triantafyllopoulos is a senior student at the University of Patras currently pursuing a diploma in the department of Electrical and Computer Engineering, majoring in Communications and Information Technology. He is a student member of the IEEE Signal Processing Society. His main academic interests are signal processing, machine learning and computer science. His diploma thesis is on real-time dynamic simulations in augmented reality scenes.

Anastasis Foliadis is a senior student at the University of Patras currently pursuing a diploma in the department of Electrical and Computer Engineering, majoring in Electronics and Computer Engineering. He is a student member of the IEEE Signal Processing, Solid State devices and Electron Devices Societies. His academic interests include signal processing, computer science, embedded systems and solid state electronics.

George Roustas is a senior student at the University of Patras currently pursuing a diploma in the department of Electrical and Computer Engineering, majoring in Electronics and Computer Engineering. He is a student member of the IEEE Signal Processing Society. His academic interests include embedded systems, analog circuits and systems, signal processing and computer science.

Ioannis Krillis is a senior student at University of Patras currently pursuing a diploma in the department of Electrical and Computer Engineering, majoring in Systems of automatic Control. His academic interests include signal processing, image processing, embedded systems and machine learning. His diploma thesis is on HDR image compression and processing.

Fanouria Athanasiou is a senior student at the University of Patras and she is currently pursuing her diploma in Electrical and Computer Engineering, majoring in Communications and Information Theory. She is a student member of the IEEE Signal Processing, Computer and Communications Societies. Her academic interests include signal processing and computer science, with emphasis on communication networks.

Maria Papaioannou is undergraduate student in the Faculty of Electrical and Computer Engineering, University of Patras, Greece. Her research interests include signal processing and nanotechnology. She is currently pursuing her thesis in the domain of molecular communication nanonetworks. She studied for a semester in the Faculty of Electronics and Telecommunications, Poznan University of Technology, Poznan, Poland as an international student.