Combining ENF Phase Discontinuity Checking and Temporal Pattern Matching for Audio Tampering Detection

Sebastian Mann, Luca Cuccovillo, Patrick Aichroth, Christian Dittmar

Fraunhofer Institute for Digital Media Technology
Ehrenbergstrasse 31
98693 Ilmenau
{sebastian.mann, luca.cuccovillo, patrick.aichroth, christian.dittmar}@idmt.fraunhofer.de

Abstract: In this paper, we present an improved approach for audio tampering detection and localization based on the Electrical Network Frequency (ENF) analysis, combining analysis of the ENF phase, and ENF temporal pattern matching: The proposed algorithm uses phase discontinuity checking to detect regions that might have been tampered, which are then matched against an ENF reference database to validate order and duration of the detected regions. Using this approach, the false-positive rate can be reduced from $\approx 55\%$ using phase analysis to $\approx 10\%$ using the combined approach, thus improving overall reliability of the tampering detection approach.

1 Introduction

Thanks to lowered cost and increased availability of devices and global networks, it is now easier than ever to record and distribute user-created audio-visual (A/V) material, always and anywhere. As a consequence to its ubiquity, such material is becoming increasingly relevant for news coverage, investigations, and many other domains, but it is crucial to be able to assess whether such content is authentic or not. Approaches for (semi-)automatic tampering detection can support this, and thus are becoming increasingly important.

In the following, we will focus on the case of tampering detection for audio recordings, which are especially easy to capture and edit, yet can convey significant meaning, and thus are an inviting target for manipulation.

As for audio tampering detection, ENF (Electrical Network Frequency) analysis represents an especially interesting approach: Due to varying conditions related to the production and consumption of electrical energy, the ENF fluctuates slightly and randomly over time rather than being fixed to an exact frequency (typically 50 Hz in Europe), and it does so commonly across the entire electrical network, typically spanning huge geographic areas. Via electromagnetic induction, many digital recordings (including e.g. battery-powered mobile device recordings near power lines) pick up the ENF, which leads to an extra low frequency component in the recorded audio signal. The ENF can thus be extracted from respective content again, via band pass filtering the recording. It can be used for various

purposes, including the following:

- Tampering detection based on ENF phase analysis: ENF fluctuations are continuous oscillations. Hence, if a continuous ENF can be extracted from an audio file, it can be assumed that the recording is original, and has not been modified. In contrast, if the ENF shows discontinuities, this can indicate possible editing of the original recording.
- ENF temporal pattern matching: It is possible to compare the ENF signature of an audio file against previously recorded ENF signatures stored in a reference database. Via pattern matching, this can be used to determine the time of the recording of the audio file.

One of the problems with tampering detection based on ENF phase discontinuity checking, however, is the amount of false-positives: Discontinuities are not only caused by tampering, but also by ENF signal faults. When this happens, the state-of-the-art methods identify the discontinuity as a tampering region border and create a false-positive classification.

In this paper, we propose a new combined approach to tampering detection, exploiting not only ENF phase analysis but also ENF matching. The idea is to use phase discontinuity checking, and then to validate the presumed tampering regions via ENF matching, thereby reducing the amount of false-positives and enhancing the overall reliability of detection. Moreover, it is possible to optimize the approach with respect to tampering localization, i.e., to localize the tampered regions.

Section 2 will provide an overview over the relevant State of the Art with respect to ENF extraction, phase analysis and temporal matching. Sections 3 and 4 will provide more details about the phase discontinuity checking algorithm and associated issues with respect to false-positives, and describe the ENF matching algorithm. Section 5 will finally describe how both can be combined, and the combined approach will then be assessed in comparison to the standalone phase analysis approach in Section 6. Section 7 concludes with an outlook on further improvements.

2 State of the Art

There are various approaches for audio authentication, including listening tests, waveform analysis, spectrum analysis, and device and environment classification methods. One important aspect are ENF-based techniques, as pointed out in [Gup12]. The ENF results in a stable tone, induced by the electromagnetic field of the power line, in the recorded signal. This happens often, both in mains-powered recording devices and battery-powered devices, if they are sufficiently close to a power line. Once extracted, ENF analysis can be used for various purposes, including phase discontinuity checking for tampering detection, and ENF matching to determine the time of a recording.

2.1 ENF extraction

The prerequisite for all further analysis is the extraction of the ENF signal. It is based mostly on either time or frequency based methods or variations of these techniques, as shown in several papers about ENF-based methods, including [KTH05], [Gri09], [San08], [Gri05] and [Coo08]. The choice between the mentioned methods depends on the application context. We focus on two cases: Extracting ENF from real-world content, and extraction from the electric power line.

Frequency-based methods are based on using the short-time Fourier transform (STFT), which operates on overlapping or adjacent frames of the audio signal. Important parameters of the transformation are the length of the Fourier transformation, the choice of the window function and the step size (which determines the overlapping). Alternative processes are Chirp-Z transform, and methods based on eigenvalue decomposition of the covariance matrix of the data.

Time-based methods measure the frequency by determining the period of the ENF oscillation (often via zero crossing detection) and by applying different filters and interpolation methods. Harmonics of the ENF-fundamental frequency are a further method to determine the ENF, but this is difficult to apply in the case of speech content, which is expected to heavily overlap with the ENF harmonics in the frequency domain.

Both techniques have almost the same accuracy in obtaining the ENF under ideal conditions, but are complementary with respect to their application context: Time-based techniques have proven to be very useful in order to record the ENF directly from the power line. Such methods are used by power suppliers, which are obliged to keep the ENF within a given tolerance and thus need to record the ENF time history to validate that. Zero crossing methods are, however, not suitable to extract ENF from real-world speech or music audio content, since they rely on the whole noise frequency spectrum for computation. Frequency/STFT-based methods, in contrast, are suitable for this, and are commonly applied for this purpose. All ENF extraction methods need some preprocessings, like bandpass filtering and downsampling.

2.2 ENF phase discontinuity checking

An important application of the ENF is tampering detection, which exploits the phase information of the ENF, as described e.g. in [NA09] and [RAB10]: The phase information from the ENF signal is extracted and then used to detect discontinuities. Assuming that cut and paste operations within audio content affect the phase (something that is difficult to avoid even for experts), phase discontinuity checking can be used to indicate editing and tampering. However, the ENF and its phase are also subject to natural variations which can lead to discontinuities and thus to false-positives. Hence, an important goal is to distinguish effectively between tampering and variations.

2.3 ENF temporal pattern matching

The ENF can be used in order to determine the date and time of a recording by using reference ENF data to match it with the ENF of a given recording, as described in [HG09]. The mean squared error (MSE) can be used as a metric to find similarities between the extracted ENF and the reference data, but MSE is not suitable for all kinds of audio recordings. As an alternative, maximum correlation coefficients or blockwise analysis can be used.

3 Phase discontinuity checking

As discussed earlier, the ENF phase is assumed to be a stable and continuous tone superimposed with the recording. Any discontinuity within the phase can indicate a tampering point, since it is almost impossible to merge two audio parts so that their ENF are blending into each other in a mathematically continuous way. To obtain the phase of the extracted ENF and thus find tampered regions, a discrete Fourier transform (DFT) of the given and windowed ENF signal $x(n)$ is used, where

$$X\left(k\right) = DFT\left(x\left(n\right)\right),$$

as proposed in [RAB10]. The phase is then simply obtained by getting the argument of $X(k)$:

$$\Phi_{ENF} = \measuredangle\left(X\left(k_{peak}\right)\right),$$

where

$$k_{peak} := \arg\max_{k \in K} |X(k)|,$$

and K is the number of frequency bins. The obtained phase of a tampered file, where a segment of another recording was spliced into the original one, is shown in Figure 1.

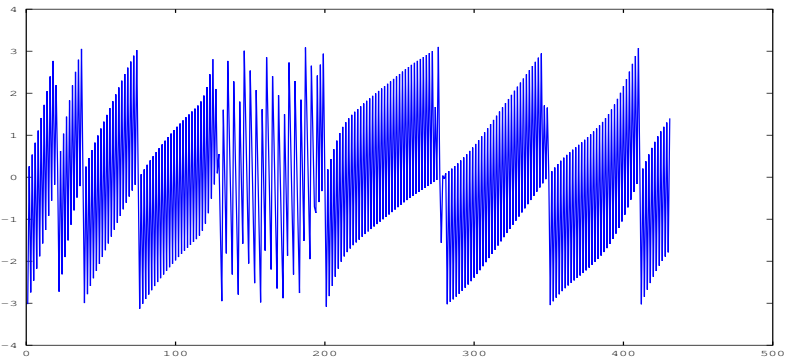

Figure 1: Visualization of the ENF phase

To detect the tampering, we locate phase discontinuities, defined as:

$$x\left(n\right) - x\left(n-1\right) > thr, \forall n = 1, \ldots, N,$$

where N is the number of samples and thr is a chosen threshold. Each detected phase discontinuity frame is marked as a suspicious point and all of these points are analyzed, because some of them only occur due to phase transitions. When several suspicious points are detected and they are all related to the same discontinuity, it is possible to aggregate them into a single suspicious frame index, i.e., the index that maximizes the variation of the mean value of the ENF phase in a surrounding analysis window. In Figure 2, the phase transitions are visible as peaks and the tampered interval has a lower phase than all other ones.

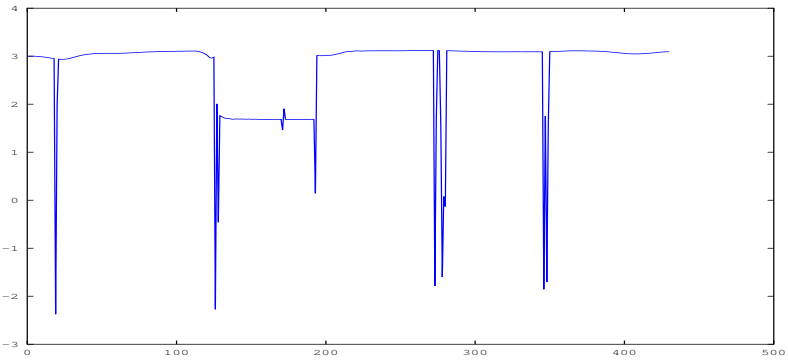

Figure 2: Visualization of suspicious frames and intervals

By providing such suspicious frames and segments, it is then possible to localize the tampering. If a segment of another audio file was spliced into the original one, we can detect the start and end frame of the spliced segment. As can be seen in Figure 2, however, discontinuities do not only appear in tampered, but also in non-tampered segments. They are caused by technical problems within the power grid, e.g. sudden peaks in energy consumption, or by ENF extraction errors. Whenever this happens, ENF phase discontinuity incorrectly classifies the segment as being tampered, leading to a high number of false positive detections. To tackle this problem, we propose the addition of ENF matching, which is described in the following Section.

4 ENF temporal pattern matching

ENF temporal pattern matching basically compares an extracted ENF against a reference ENF database in order to determine the time of a recording, or to verify an alleged recording time. For this purpose, a huge database of archived ENF data is necessary. For our research, we received such ENF reference data from a German energy network operator for evaluation purposes. The reference data is sampled with 1 Hz and spans a period of several days, including the timespan the test recordings were made in. An ENF time series is extracted from the original audio file, which can be matched with the reference data. This method needs an interval, which is long enough to allow a robust matching.

We use the MSE method for matching extracted ENF against reference data as proposed in [HG09]. A smaller MSE indicates that the two vectors are more alike, and it is defined as:

$$E = \left(\frac{1}{L} \sum_{i=1}^{L} (x_i - y_i)^2 \right), \tag{1}$$

where L is the number of samples of the two vectors x and y. When matching an extracted ENF against a database, there is typically one shorter vector (the recorded ENF r) and one longer vector (the database vector d). The approach of Huijbregtse is to calculate a number of MSE's while sliding the recorded ENF vector across the longer database vector sample by sample.

$$e[k] = \sum_{i=1}^{R} (r[i] - d[i+k-1])^2, \tag{2}$$

where R is the length of the recorded ENF. The index k runs from 1 to $D-R+1$, where D is the length of the database vector. Then, the minimum value of the vector e determines the best match between the recorded and the database vector. Figure 3 shows the extracted ENF of the given audio file (red), overlayed with the curve of the corresponding ENF database (blue) and the match between the database and the test file ENF (green).

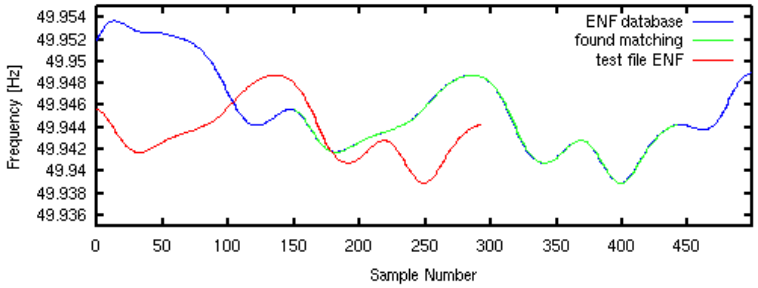

Figure 3: Visualization of the temporal matching of the ENF with the database

This method works properly for audio files where ENF is detectable over a long period of time. However, in many cases, ENF is only detectable within short time intervals of silence, where the ENF is not overlaid by noise. Hence, content segments containing detectable ENF are separated from each other by parts without detectable ENF.

5 Combination of ENF Phase Discontinuity Checking and Temporal Matching

The main goal of the combination of ENF phase discontinuity checking and temporal matching is the improvement of the accuracy by reducing the false-positive rate. This can be achieved by first searching for suspicious frames, using the phase discontinuity

checking, resulting in discontinuities which are either caused by tampering, or by ENF faults. Then, to distinguish between the two types of discontinuities by checking whether the corresponding Section in the ENF database is marked as valid or not. If it is valid, the interval is considered to be a possible tampering.

This approach, however relies on a previous knowledge of the time of the recording, which is not available to an analyst.

To address this, Figure 4 describes the refined process for the combined approach:

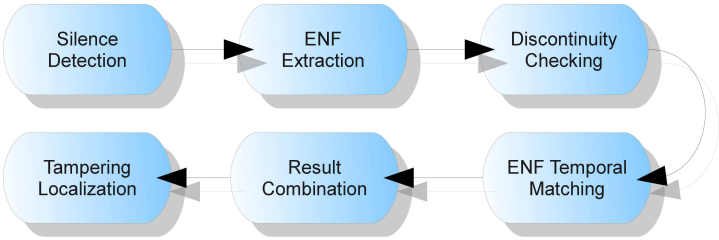

Figure 4: Process flow of the proposed combined method

1. As a first step, silent parts within the audio file are identified, using a silence detector.
2. Then, the presence of the ENF is visually detected within the silent parts and after that automatically extracted as a time series (see Figure 5, which shows the visualization of the ENF of a given audio file).
3. Subsequently, tampering detection via phase discontinuity checking of the extracted ENF is performed, resulting in suspicious frames and intervals (see Figure 2).
4. Afterwards, the aforementioned ENF matching algorithm is performed on the detected intervals (see Figure 6, which shows the outcome of the MSE-based ENF temporal matching algorithm).

Using this approach, it is possible to check whether the aligned intervals are matching the corresponding Section from the database, or whether one or more of them are matching a different database entry. If the arrangement or duration of the intervals detected by the ENF discontinuity check are not compatible with those computed via ENF temporal matching, the audio file is classified as tampered.

This procedure can reduce the number of tampering detection false-positives: The ENF discontinuity check by itself is not able to distinguish between ENF faults and discontinuities due to a real tampering. The ENF temporal matching algorithm, on the other hand, can accomplish this task, thus reducing the number of false-positives.

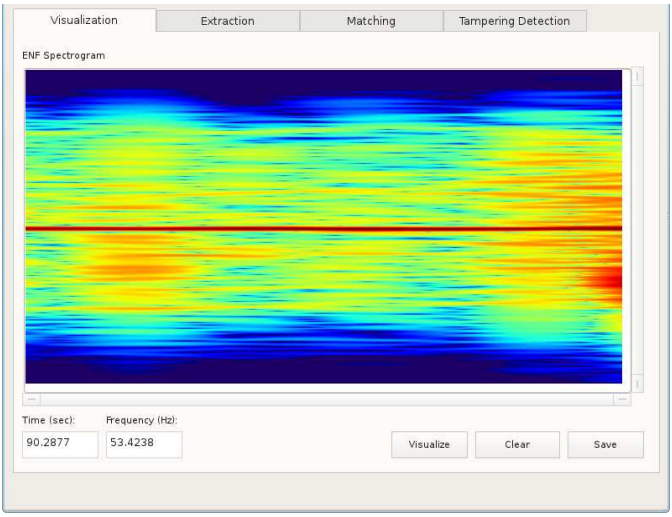

Figure 5: Visualization of the ENF

6 Evaluation

Considering the fact that ENF is present in most mains-powered devices, a main-powered laptop Dell Latitude D630 was used to acquire the test recordings. The recordings include different types of content, as reported in Table 1. The original recordings used during

Table 1: Test content for evaluation

Label	Content
1	Speech content (male voice)
2	Speech content (female voice)
3	Speech content (dialogue, male and female voice)
4	Speech and music content replayed by a loudspeaker
5	Silent content with external noise source

the test content generation share a similar feature, i.e., the ENF is continuous and clearly visible. The testing involved a total of **59** original recordings and **353** tampered ones, for a total of **412** tampering detection tasks. If a previous knowledge of the supposed time of the recording is given, it is possible to enhance the performance, by constraining the interval of the reference data to an assumed time interval. If not, the matching algorithm must consider the complete database. The baseline performance was assessed by using only the ENF phase discontinuity checking as a stand-alone algorithm. Afterwards, the proposed combined approach of ENF phase discontinuity checking and ENF temporal matching was used. The detailed results are given in Table 2 for the localization of the tampered region and in Table 3 for a binary classification (tampered vs. non-tampered).

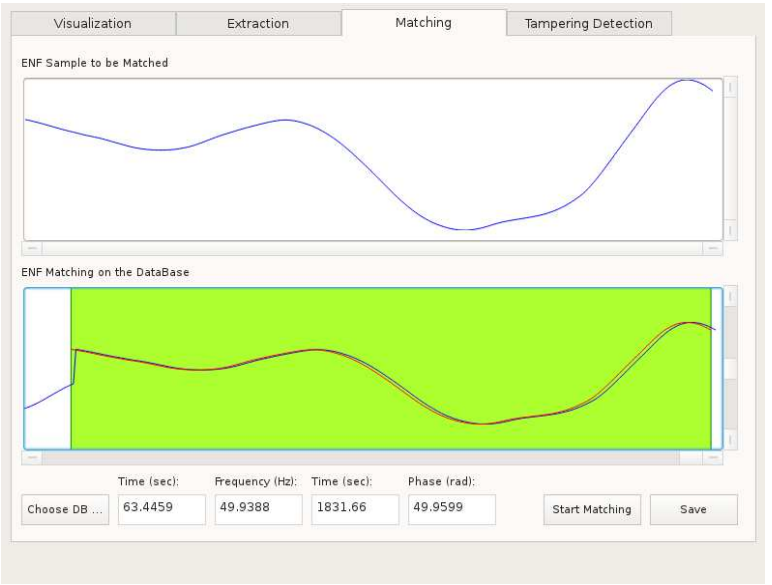

Figure 6: Visualization of the outcome of the ENF temporal matching algorithm

Table 2: Interval Detection Results

Algorithm	Outcome (#)				Statistics (%)			
	TP	FP	TN	FN	Precision	Recall	FP rate	Accuracy
Phase	499	331	52	207	60.12	70.68	86.42	50.60
Combined	486	189	150	220	72.00	68.84	55.75	60.86

Table 3: Tampering Detection Results

Algorithm	Outcome (#)				Statistics (%)			
	TP	FP	TN	FN	Precision	Recall	FP rate	Accuracy
Phase	271	33	26	82	89.14	76.77	55.93	72.09
Combined	270	6	53	83	97.83	76.49	10.17	78.40

In Table 2, true positive indicates that a tampered region was correctly identified as such. Despite the accuracy being far from being satisfying (50.60% - 60.86%), it is important to note that the the number of false positives suddenly drops with the combination of ENF phase discontinuity checking and ENF temporal matching.
In Table 3, true positive indicates that a tampered audio file was correctly classified as tampered. We can see that both the ENF phase discontinuity checking as a stand-alone and the proposed algorithm have nearly the same recall (76.77% - 76.49%), but the precision

of the combined approach is much higher, i.e., 97.83% against 89.14%. This also reflects on the accuracy, that rises from the 72.09% of the ENF phase discontinuity checking up to 78.40%.
These results show that the combined approach has nearly the same detection capability as the baseline, but that a detection performed by the combined approach is much more reliable than one achieved by phase discontinuity check. Moreover, the initial assumption about the reduced number of tampering detection false positives was confirmed by the experimental results: The false positive rate, defined as

$$\frac{FP}{FP+TN},$$

with the combined approach drops from 55.93% down to 10.17%.

7 Conclusion and Future Work

In this paper, we proposed a new combined approach for tampering detection, by combining a state of the art method (ENF phase discontinuity checking) with another method that was not designed for tampering detection (ENF temporal matching), thereby improving the approach, but also making it possible to locate tampering within the audio file. The proposed method exhibits an accuracy that is significantly higher than the one of phase discontinuity checking, even if still insufficient with 78.40%. The quality of the tampering detection, however, is improved significantly, with the false positive rate dropping from 55.93% to 10.17%, and the precision increasing from 89.14% to 97.83% when comparing ENF phase discontinuity checking standalone versus the new combined approach.

Further enhancements are possible: Many possible improvements for the ENF temporal matching itself can be applied, in order to further reduce the number of occurrences of false positives, e.g., refinement of the matching algorithm by using other matching methods, introduction of robust ENF database recording or robust ENF extraction from an audio file. Furthermore, any improvement of the ENF phase discontinuity checking would result in an higher recall, thus improving the overall accuracy. Finally, it also seems possible to combine ENF methods with completely different tampering detection approaches.

8 Acknowledgments

This research has been partially funded by the EU project REWIND. The project acknowledges the financial support of the Future and Emerging Technologies (FET) Programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 268478.

References

[Coo08] A.J. Cooper. The Electric Network Frequency (ENF) as an Aid to Authenticating Forensic Digital Audio Recordings an Automated Approach. *In Audio Engineering Society Conference*, 2008.

[Gri05] C. Grigoras. Digital Audio Recording Analysis: The Electric Network Frequency (enf) Criterion. *International Journal of Speech Language and the Law*, 12(1):1350 – 1771, 2005.

[Gri09] C. Grigoras. Applications of Enf Analysis in Forensic Authentication of Digital Audio and Video Recordings. *Journal of the Audio Engineering Society*, 57(9):643 – 661, September 2009.

[Gup12] Kuo Gupta, Cho. Current Developments and Future Trends in Audio Authentication. *Multimedia in Forensics, Security and Intelligence*, pages 50 – 59, 2012.

[HG09] M. Huijbregtse and Z. Geradts. Using the ENF Criterion for Determining the Time of Recording of Short Digital Audio Recordings. *In Proc. 3Rd Intl. Workshop Computational Forensics*, 2009.

[KTH05] M. Kajstura, A. Trawinska, and J. Hebenstreit. Application of the Electrical Network Frequency (ENF) Criterion: A Case of a Digital Recording. *Forensic Science International*, 155(2 - 3):165 – 717, 2005.

[NA09] D. Nicolalde and J. Apolinario. Evaluating Digital Audio Authenticity with Spectral Distances and ENF Phase Change. *In Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Processing*, 2009.

[RAB10] D.P.N. Rodriguez, J. Apolinario, and L.W.P. Biscainho. Audio Authenticity: Detecting ENF Discontinuity with High Precision Phase Analysis. *IEEE Trans. Information Forensics and Security*, 5(3):534 – 543, 2010.

[San08] R.W. Sanders. Digital Audio Authenticity Using the Electric Network Frequency. *Audio Engineering Society Conference*, 2008.