ENF PHASE DISCONTINUITY DETECTION BASED ON MULTI-HARMONICS ANALYSIS

Valentin A. NIȚĂ[1], Amelia CIOBANU[2], Robert Al. DOBRE[3],
Cristian NEGRESCU[4], Dumitru STANOMIR[5], Radu O. PREDA[6]

In this paper a multi-harmonic electric network frequency (ENF) phase discontinuity detection method is presented, used for detecting possible editing points in an audio recording. After we reveal the importance of a multi harmonic approach through a brief analysis on how the hum noise contamination manifests in an audio recording, we propose a system for tracking the ENF harmonics by performing a high precision frequency analysis only over the silence parts in an audio recording. Based on an energy related criterion we detect possible phase discontinuities through a wavelet decomposition algorithm and determine the positions of possible editing points. The detected positions are compared across harmonics and a final decision is delivered. Test results show that the proposed system is able to correctly identify editing points that could not be detected if only the first harmonic was monitored.

Keywords: ENF, multi-harmonics, phase discontinuity

1. Introduction

The electric network frequency (ENF) is the power system AC frequency which, from a theoretical point of view, is a pure sinusoidal signal of 50Hz (Europe and almost all Asia) or 60Hz (North and South America). In [1] it is proven that in reality the power grid frequency varies around 50/60Hz, the variation is not periodic and it is the same for large geographical areas. These characteristics make ENF a good candidate for time stamping an audio recorded with an equipment connected to the AC power grid.

---

[1] Dept.of Telecommunication, University POLITEHNICA of Bucharest, Romania, e-mail: vnita@elcom.pub.ro
[2] Dept.of Telecommunication, University POLITEHNICA of Bucharest, Romania, e-mail: aciobanu@elcom.pub.ro
[3] Dept.of Telecommunication, University POLITEHNICA of Bucharest, Romania, e-mail: rdobre@elcom.pub.ro
[4] Dept.of Telecommunication, University POLITEHNICA of Bucharest, Romania, e-mail: negrescu@elcom.pub.ro
[5] Dept.of Telecommunication, University POLITEHNICA of Bucharest, Romania, e-mail: dumitru.stanomir@elcom.pub.ro
[6] Dept.of Telecommunication, University POLITEHNICA of Bucharest, Romania, e-mail: radu@comm.pub.ro

The time stamping of the audio signal is possible if an ENF database is available for matching the extracted ENF with the database entries. However, as [2] reveals, in order to obtain good results, the duration of the investigated audio signal has to be at least one hour long. This implies that small changes performed over the audio recording, like word insertion or word cutting, cannot be observed based on the ENF pattern matching. To increase the robustness of the ENF method even for shorter audio recordings, a maximum correlation coefficient-based matching procedure is introduced in [2].

ENF has another particularity that can be explored in situations which deal with short audio signals: it is present on the entire audio recording. Based on this particularity, in [3] it is described a method for finding possible editing points in an audio recording, by monitoring the ENF phase discontinuity. However, analyzing only the first ENF harmonic is not a full proof solution, because the first harmonic can be easily forged since it does not spectrally overlap with the recorded audio signal. Having this in mind, a better solution is to look for discontinuities of phase, not only on the first harmonic of the ENF, but on multiple harmonics [4]. A multi-harmonic analysis is much harder to forge without leaving artifacts because the ENF harmonics can often be placed in the vicinity of salient speech components, such as the fundamental frequency. If such situation should occur, the forgery could be easily detected using one (or more) of the methods presented in [5].

Preserving the idea of finding possible editing points through harmonic ENF analysis (see [4]), in this paper we propose to concentrate the analysis only over the silence segments found in an audio recording. This approach has several advantages. First, it considerably reduces the amount of processed data without diminishing the accuracy of the analysis procedure in finding possible editing points. This is possible because in general forgeries are operated between moments of speech, i.e. in areas where the signal's level is low enough so that a modification is inaudible. Second, the presence of the ENF harmonics is detected in a simple manner using a common spectral analysis of the silence segments. Thus, the need for additional filtering, as in [4], is eliminated.

The framework of the proposed approach starts with a voice activity detection (VAD) procedure, followed by a high precision spectral analysis at the end of which the parameters of the ENF harmonics are detected. Among these parameters, the phases of the ENF harmonics contain important information regarding the existence of possible editing points. Consequently, the track of each ENF harmonic is tested for discontinuities. We propose to complete this task, in an automatic manner, by using a Haar wavelet decomposition algorithm.

The paper is organized as follows. In Section 2 we analyze the situations in which the ENF and its harmonics are prone to occur in a digital recorded audio signal due to hum noise contamination. In addition to that, using our own

designed power supply, we reveal the spectral characteristics of the ENF harmonics. In Section 3 we present the framework of the proposed system for determining insertion and/or cutting points in an audio recording and we provide a simpler solution for deciding if the ENF signal is present or not in the analyzed audio recording. Section 4 briefly describes the spectral solution adopted for a precise finding of the ENF harmonics' parameters, namely the frequencies and phases of the harmonics. Section 5 covers the wavelet decomposition method for determining the phase discontinuities in ENF multi-harmonics. Finally, the last section is reserved for conclusions and future work.

2. General considerations regarding the ENF harmonics

The electrical network signal is often related to the hum noise present in most of the audio recordings. In this section we reveal the characteristics of the ENF signal extracted from the hum noise produced by a power supply, specially designed by us to mimic the real situations.

The common causes which lead to the appearance of the hum noise are bad filtered power supplies or ground loops. An economically designed power supply represents a common source of hum noise. The DC operating point for the output of the audio system is considered to be 0V, which is a typical situation. These equipment use dual power supplies to obtain the positive and negative DC rails. These are not identical, they will not present the same voltage ripple because the filtering circuits are not perfectly matched and the current drawn from one rail is not necessary equal to the current drawn from the other one. It can be considered that one rail has no ripple and the other contains the equivalent ripple for both the rails. In the presented situation it was considered that the positive DC rail contains the equivalent ripple voltage. The only difference between the case treated in this paper and the real situation is the DC component, however it is unimportant in the context of our analysis.
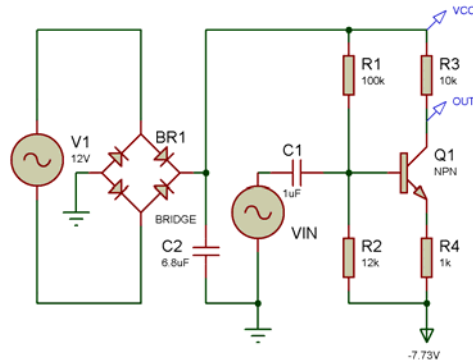

Fig. 1. The schematic used for DC power supply simulation.

In Fig. 1 we propose a typical transistor based amplifying circuit for simulating the hum noise contamination. By using different values for the capacitor C2 the quality of the DC filtering is modified, a larger capacitor determining a better filtering, eliminating the harmonics contained in the full-wave rectified sine wave obtained after the diode bridge and keeping only the needed DC component. If the capacity is large enough, the ripple voltage will be negligible and the output signal will be an amplified replica of the input signal centered on the DC operating point, after the transitory regime has ended. This behavior can be observed in Fig. 2.
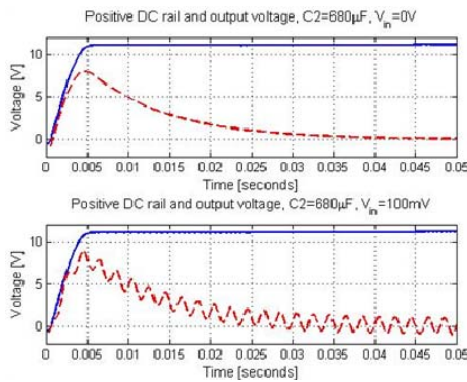

Fig. 2. Simulation results for well filtered power supply. Continuous line – positive DC rail, dashed line – output signal
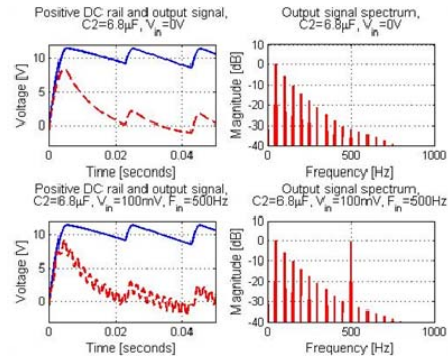

Fig. 3. Simulation results for badly designed power supply. Continuous line – positive DC rail, dashed line – output signal.

If the capacity of C2 is small then the output will be affected by hum noise permanently. The hum noise is additive to the output signal, as it can be observed in Fig. 3. The spectral content of the output signal is also presented, demonstrating that multi-harmonics analysis should be performed to extract information about the electrical network signal.

Another common situation that determines the hum noise to enter a system is represented by ground loops. When different audio equipment are interconnected using an unbalanced connection, if they are plugged in different outlets, it is very probably that there will be a voltage between their ground connections. A current will flow through the shield of the cable used for the connection. The voltage difference will be treated as an input signal by the receiver, determining hum noise to enter the system.

3. The multi-harmonics analysis

The analysis in Section 2 indicates the strong presence of the ENF harmonics in the electric network signal, therefore concedable information

regarding the characteristics of the electric network signal is carried by these components as well.

We propose to determine if an ENF harmonic is present by performing a spectral analysis only over the silence segments found in an audio recording. This solution on the one hand reduces the overall computation effort, and the other hand eliminates the interference of the relevant speech components in the analysis of the ENF. Moreover, the solution is in agreement with the real situations in which audio forgery is produced. Namely, in general an audio editing is operated over speech segments with very low acoustic level (e.g. silence) so that the modification is not audible.

Taking into consideration the observations above, we start the multi harmonic analysis proposed in this paper with a voice activity detection procedure – VAD (see Fig. 4). In our implementation we used the solution provided in [6].
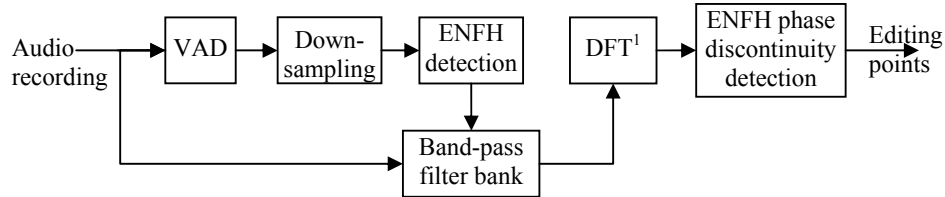

Fig. 4. The block diagram of the proposed multi-harmonic analysis.

The VAD routine is followed by a downsampling procedure. The reason behind this approach is directly connected to the analysis performed in Section 2 which revealed that the most relevant ENF harmonics appear in the spectrum for frequencies less than 500Hz. Thus we reduced the original sampling frequency to 1000Hz. As a side effect, the computational effort was also diminished. Next, for every detected silence segment in the audio recording, with a duration greater than 0.5s, we decide if an ENF harmonic (ENFH) is present by employing a simple discrete Fourier transform (block *ENFH detection* in Fig. 4) in conjunction with an energy criterion, detailed further in the current section. If a positive decision occurs (an ENF harmonic is present) then the entire signal is passed through a band-pass filter bank in order to isolate the detected ENF harmonics. For each band-pass filtered signal a high precision frequency analysis based on the signal's derivative (further referred to as DFT^1 [7]) is performed in order to extract the frequency and phase information of the ENF harmonic. The DFT^1 analysis is applied on analysis frames with a duration of 0.1s. The hop size of the analysis frame is chosen equal to the inverse of the ENF (0.02s) so that the initial phases of the ENF harmonics are kept constant for all the analyzed frames. This approach helps to obtain, in an ideal case, a flat variation for the ENF phase estimates, therefore discontinuities are more easily spotted. Finally, the last block in Fig. 4 computes the phase tracks of the analyzed ENF harmonics and searches for any

discontinuities that may appear in these tracks. The existence of a discontinuity indicates that a forgery took place in the audio recording and its position is determined by the position of the phase discontinuity.

An important step in the multi harmonic analysis proposed in this paper is to decide whether ENF harmonics are present or not in the analyzed audio recording. Without a solid decision in this direction, the phase estimates will be unreliable and no analysis can be performed on the phase tracks.

A first attempt to solve this problem was based on the eigenvalues of the covariance matrix associated to band-pass filtered versions of the analyzed signal, performed for different bandwidths. The decision criterion, inspired from the one in [4], involved the ratio between the eigenvalues of the output from a large bandwidth filter (20Hz band) and a narrow one (2Hz band). The proposed method was tested for 16000 signals with or without ENF harmonics, at different signal to noise ratios (SNRs), and for different noise power. The results revealed a detection rate close to 100% for positive SNRs, however the computational complexity of the method determined us to develop a much simpler solution but with similar outcome.

Consequently, for each silence segment delivered by the VAD routine we compute the amplitude spectrum using the fast Fourier transform (FFT). Then the mean magnitude spectrum is obtained by averaging all the silence spectra, resulting the lower threshold in Fig.5, denoted with L_0.
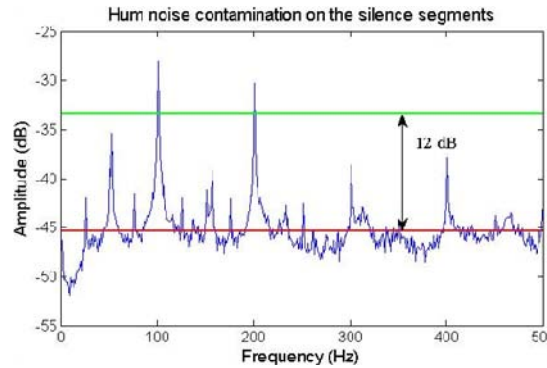

Fig. 5. DFT analysis on the silence segments.

The algorithm decides that a harmonic is present, if its magnitude exceeds a certain threshold resulted from the estimation errors obtained in the high frequency and phase analysis described in Section 4. After numerous experiments we found that a good threshold is:

$$L = L_0 + 12\text{dB} .$$ (1)

For peaks under the threshold, L the phase estimation errors cannot be controlled and lead to too many false detections, i.e. significant discontinuities

appear in the phase tracks, comparable with the discontinuities introduced by actual forgeries.

4. High precision frequency and phase analysis

For an ENF analysis a high precision frequency estimation algorithm is required, because frequency variations of ±0.1Hz around the nominal value 50/60Hz are common for our application and it is very important to detect them correctly. In order to obtain good precision (e.g. 0.001Hz) with a classical FFT, the analyzed frame has to be either long enough or zero padded so that its length reaches approximately 10^6 samples. Since the analyzed signal is non-stationary, choosing long analysis frames is not an option, therefore zero padding can be a solution, however this will result in a high computational cost, especially if there are hours of signal to be checked for audio authenticity.

In [7] a high precision frequency analysis, but with a low computational complexity is presented. The method assumes that the analyzed signal is a single tonal signal and is based on the signal's derivative.

For signals containing more than one tonal component, as it is the case in our application, in order to apply the method described above, we filter out each harmonic whose level is above the threshold L using a very narrow band-pass filter. Then for each filtered signal we apply the DFT[1] method.

In the last part of this section we present the performances of the DFT[1] method applied in the context of the ENF harmonic analysis.

The first scenario consisted of 10000 test signals for which the frequency and phase of the tonal components was estimated using the DFT[1] method. Each signal contained components placed on the first 8 harmonics, with arbitrary initial phase and a fundamental frequency that varied uniformly in the interval $[49,51]$Hz. The parameters N (the length of the analysis window) and N_{FFT} (the FFT length) were varied, in order to find the best configuration. The sampling frequency was 1000Hz. Table 1 presents the estimates of the frequencies and phases of the tonal components, expressed in terms of mean absolute error (e_f for frequency and e_p for phase) and standard deviation (σ_f and σ_p).

It can be observed that the configuration with $N = 100$ and $N_{FFT} = 8000$ is a good compromise between the computational cost, the frame's length and the algorithm's capacity to capture phase changes. In the second scenario we tested the algorithm's performances with respect to the signal to noise ratio (SNR). We used the same set of test signals as in the first scenario for which we varied the SNR between 0 and 50dB. Regarding the length of the analysis frame and the number of FFT points, the configuration determined in the first scenario was used ($N = 100$ and $N_{FFT} = 8000$).

Table 1

Frequency and phase estimation results for different frame lengths and FFT lengths

N	N_{FFT}	Frequency		Phase	
		e_f [Hz]	σ_f [Hz]	e_p [deg]	σ_p [deg]
60	1200	0.126	0.297	1.274	2.911
	2000	0.108	0.285	1.086	2.783
	8000	0.083	0.280	0.828	2.726
100	1200	0.062	0.075	1.105	1.336
	2000	0.043	0.053	0.765	0.933
	8000	0.016	0.031	0.286	0.542
200	1200	0.061	0.071	2.203	2.563
	2000	0.039	0.043	1.412	1.572
	8000	0.010	0.012	0.371	0.428

The results show that this method can be used for SNRs as low as 10dB, or even 0dB if the phase discontinuity is considerably bigger than 15.4 degrees, the mean error phase estimation.

5. Phase discontinuity detection

As the test results in Section 4 reveal, depending on the SNR of the analyzed signal and on the frame length used, the estimated phase can have a mean error up to 5 degrees with up to 3 degrees standard deviation. Due to this small variation around the mean error, we can observe in Fig. 6, for the second and forth harmonic, that the variation of the estimated phase of the ENF signal can be considered linear on long term, but on short terms it presents small fluctuations. Therefore we need a robust algorithm to differentiate between the variations caused by estimation errors and the variations caused by modifications made on purpose to the recorded audio signal. The latter ones generally present larger variations.

Fig. 6 illustrates an example of the phase tracks corresponding to harmonics above the detection threshold *L* (the second and forth ENF harmonic as results from Fig. 5) and below *L* (the first harmonic or ENF), extracted from an edited signal in which an insertion and a cutting was performed around the time moments 7s and 24s respectively. The phase track of the first harmonic was chosen (although is not indicated as valid) to emphasize the side effects of an analysis based on invalid harmonics. If no audio forgery is present in the original signal, then the phase tracks of the analyzed harmonics should be relatively constant (provided that the analysis hop size is equal to an integer multiple of the inverse of the fundamental period). A modification applied to the audio recording will lead to discontinuity jumps in the phase evolution. In the example considered here, on each phase track three discontinuity points should be visible, two from the 3s insertion and one from the cutting operation.

The results obtained with the proposed system show that on each track in Fig. 6 at least one significant discontinuity can be observed, which corresponds to an insertion point. If only the first harmonic (the invalid one) were to be analyzed, then the cutting point could not have been identified. When taking into account the second and forth harmonic phase tracks (the valid ones) we can easily observe all the discontinuity points corresponding to the real forgery points.

The last step in our algorithm is to automatically detect the discontinuity points from the phase tracks. The discontinuity detection algorithm should be robust to noise, therefore we considered the solution presented in [8] because it provides good results in noisy environments and has low computational costs. The method is based on the wavelet decomposition, described for 2D signals [8], therefore we adapted the solution for our 1D signals. In the end our particular solution relies on the 4th level approximation wavelet decomposition, using Haar wavelet family.
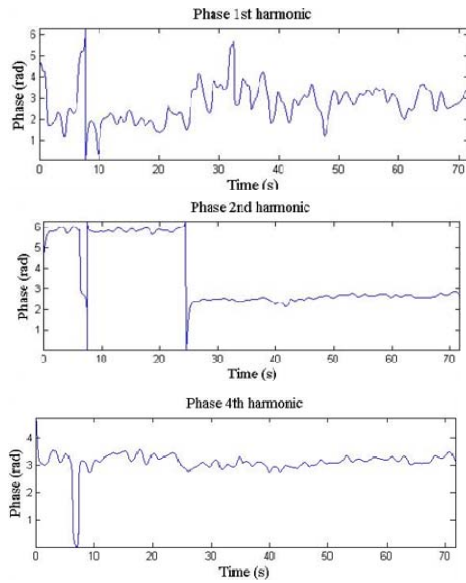

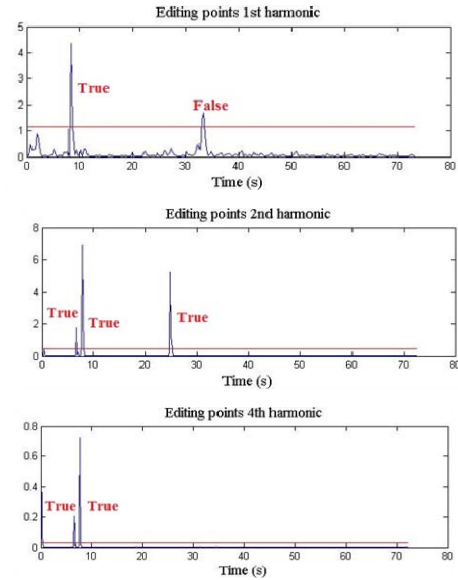Fig. 6. Phase estimation of a real recorded audio signal of the ENF's first, second and forth harmonic

Fig. 7. Phase discontinuity detection using a forth level Haar wavelet decomposition, for the first, second and forth harmonic

We consider that a discontinuity is present if we detect peaks with amplitude greater than five times the mean of the used decomposition. The threshold emerged from the experimental tests we performed on 200 signals, with an average duration of 15s. Its value was chosen high enough in order to significantly reduce the false alarm rate, but at the same time, small enough to capture the audio editing points.

In Fig. 7 we illustrated the results of the discontinuity detection algorithm for the example considered in Fig. 6 using a HAAR wavelet decomposition on the phase estimation data for the first, second and forth ENF harmonic, for a test signal contaminated by the hum noise as in Section 2. We computed the absolute value of the fourth level approximation wavelet decomposition. The peaks over the threshold indicated with the straight line in Fig. 7, show the existence of the editing points.

We can observe that for the second and forth harmonic the discontinuity algorithm we use is able to find all the editing points, and moreover there are no false alarms. Regarding the first ENF harmonic, the proposed algorithm detects only one out three editing points plus a false alarm. Another interesting result that can be observed in Fig. 7 is that the phase of the 4th harmonic does not capture the last editing point, because the discontinuity introduced for this harmonic is comparable with the phase estimation errors. However, if the discontinuities in all the valid harmonics are evaluated, all the editing points can be indicated.

In order to validate the proposed audio forgery detection system based on the multi-harmonic analysis we used 72 test signals sampled at 8000Hz, each signal with an average duration of 15s. The test signals were speech excerpts randomly extracted from studio recordings over which we added hum noise produced with the schematics in Fig. 1. Half of the test signals contained one editing point, while the other half was left intact.

According to the diagram block in Fig. 4, after the VAD routine, the signals were resampled to 1000Hz. For the DFT[1] analysis we used analysis frames with a duration of 0.1s and a hop size of 0.02s. The particular choice of the hop size (the inverse of the nominal ENF) ensures a relatively flat variation of the phase estimates, if no editing is performed on the analyzed signals.

Table 2

The performances of the proposed system

Valid harmonics considered in the final decision	Detection [%]	False alarm [%]
50 Hz	76.67	1.38
50 Hz, 250 Hz and 350 Hz	44.44	0
At least two out of: 50 Hz, 250 Hz or 350 Hz	83.33	0

Table 2 summarizes the test results from our experiments, in different situations. In all the analyzed frames, the number of valid harmonics was 3 (the valid harmonics were placed on 50Hz, 250Hz and 350Hz). If we consider the phase tracks of the first harmonic, then the system is able to correctly identify the existence and the location of the editing point in only 76.67% of the cases and in 1.38% of the cases the system detects a false discontinuity. If we monitor all the three harmonics and we validate a discontinuity based on its visibility on all the

phase tracks we obtain 0% false alarm but the detection drops to 44.44%. When we validate the discontinuity based on at least two out of the three phase tracks, the system's performances increase, reaching a maximum of 83.3% and the false alarm remains 0%.

Regarding the false alarm percentage, the test results show that it is less than 2%, which means that the possibility to erroneously consider that an editing took place when in fact it did not, it is extremely low. Although neither the misdetection, nor the false alarm are desired in any system, in the context of audio forensic, a system with low false alarm rate is preferred over a system with low misdetection rate, but with high false alarm because it is always better to exonerate a person due to lack of evidence, rather than accusing her/him of something she/he did not commit.

Other comparisons between the performances of our solution and others reported in the literature (e.g. [4]) were not possible due to different testing conditions, namely different audio databases.

6. Conclusions

In this paper we proposed an audio forgery detection system based on an ENF multi-harmonic analysis. The novelty in our system relates to several aspects. First, we propose to decide if higher ENF harmonics are present in the analyzed signal by inspecting only the silence segments. This solution is in perfect agreement with the manner in which most audio editing is performed and reduces the overall computational cost. Second, we simplified the decision algorithm by introducing an energy criterion based on which we identify the valid harmonics. Reliable phase estimates extracted with a high precision analysis method are further used to obtain the phase tracks corresponding to the valid harmonics. In the final part of the system we proposed to make the final decision regarding the existence of editing points by searching for the possible phase discontinuities in all the valid harmonics. This task is completed with the help of a fourth level approximation wavelet decomposition for all the valid harmonics. When considering the discontinuities over at least two ENF harmonics, the system is able to produce 83.33% correct detection rate with 0% false alarms.

Consequently, analyzing all the ENF harmonics validated by the proposed energy criterion applied only for the silence parts of the audio recording, we obtained a forensic tool suited for audio forgery detection, with an extremely low probability of false alarms.

In the end a final remark regarding the system's performances is in order. The test signals used in our experiments contained hum noise artificially produced in the manner described in Section 2, for which the higher harmonics had a strong presence. Preliminary test results showed that in real practice, if the higher

harmonics spectrally overlap with the speech components, then the phase estimation is not accurate and the system's performances decrease. We found that if the magnitude of the ENF harmonic is not at least 12dB over the speech component placed in the close vicinity of the same ENF harmonic, then the phase estimates are not reliable, and the detection is impaired especially for female voices where the fundamental frequency is higher.

In the near future we intend to extend the tests in order to cover more real case scenarios where the electric network is affected by power spikes or the recorded signal is clipped on several portions and determine methods for eliminating false alarms in these situations.

Acknowledgement

R E F E R E N C E S

[1]   *C. Grigoras*, Digital audio recording analysis: the electric network frequency (ENF) criterion, Int. J. Speech Language Law, **vol. 12**, no. 1, pp. 63–76, 2005.

[2]   *M. Huijbregtse, and Z. Geradts*, Using the ENF Criterion for Determining the Time of Recording of Short Digital Audio Recordings, Computational Forensics Lecture Notes in Computer Science, **vol. 5718**, pp. 116-124, Jul./Aug. 2009.

[3]   *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, **vol. 5**, no. 3, pp. 534 – 543, September 2010.

[4]   *D.P.N. Rodriguez, J. Apolinario, and L.W.P. Biscainho*, Audio Authenticity Based on the Discontinuity of ENF Higher Harmonics, in Proc. EUSIPCO, 2013.

[5]   *R. Maher*, Audio forensic examination, IEEE Signal Processing Magazine, **vol. 26**, no. 2, pp. 84–94, March 2009.

[6]   *J. Sohn, N. S. Kim, and W. Sung*, A statistical model-based voice activity detection, IEEE Signal Processing Letters, **vol. 6**, no.1, pp. 1-3, January 1999.

[7]   *M. Desainte-Catherine and S. Marchand*, High-precision fourier analysis of sounds using signal derivatives, J. Audio Eng. Soc., **vol. 48**, no.7/8, pp. 654–667, Jul./Aug. 2000.

[8]   *E. Rufeil1, J. Gimenez, and G. Flesia* Comparison of edge detection algorithms on the undecimated wavelet transform, in Proc. CLAM, 2012.