Detection of ENF Discontinuities Using PLL for Audio Authenticity

Magdalena Fuentes, Pablo Zinemanas, Pablo Cancela
Instituto de Ingeniería Eléctrica, Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay
{mfuentes,pzinemanas,pcancela}@fing.edu.uy

José Antonio Apolinário Jr.
Military Institute of Engineering (IME)
Programs of Electrical and Defense Engineering
Rio de Janeiro, Brazil
apolin@ime.eb.br

***Abstract*—A Phase Locked Loop (PLL) based method for determining audio authenticity is proposed in this work. Assuming that the power grid signal is embedded in an audio signal, certain pre-processing techniques are applied in order to obtain the Electrical Network Frequency (ENF). A PLL is then used to estimate the time-varying phase of the ENF signal. Postprocessing is carried out so as to improve system performance. Finally, an automatic decision on the authenticity of the audio is conducted by quantifying the frequency variations of the VCO output. The performance of the proposed method is evaluated on digitally edited and original audio signals, with promising results (achieving an accuracy of 96%).**

***Keywords*—Audio authenticity, Phase-Locked Loop (PLL), Electric Network Frequency (ENF).**

I. Introduction

Due to the increment of technology availability in the recent decades, editing digital audio has became a very simple task [1]. Besides this, if the edition is carried out carefully, it is hard to determine if the audio has been modified or not, even for trained ears. In some cases (e.g. when the recording device is connected to an electrical outlet), the power grid signal is usually embedded in the recorded signals. In recent years, several works focusing on detecting electric network frequency (ENF) discontinuities for audio authentication have been carried out. Most existing research involve the Discrete Fourier Transform (DFT) or spectral distances for extracting the phase changes of the ENF [2]-[4]. In this work, we propose an alternative method based on tracking the ENF with a PLL. The method is specially tailored for the analysis of audio signals recorded with the presence of the power grid signal (50 Hz or 60 Hz).

The rest of this paper is organized as follows. The next section provides some basic background on the power grid signal. Section III outlines the proposed approach for audio authenticity. Section IV is devoted to detail the experiments used to evaluate the proposed system, its results being also presented in there. Concluding remarks are given in Section V.

II. The Electric Network Frecuency

Most of the power from a network comes from turbines that drive alternating current generators; the rotating speed of their turbines determines the ENF [5]. Standards adopted by countries worldwide are either 50 Hz or 60 Hz ENF nominal values. The system frequency will vary around the target, 50 Hz or 60 Hz, and the network operators have statutory obligations to maintain the frequency within certain limits (in the order of 0.5 Hz for developed regions [5]).

Any operating electrical equipment connected to the power grid produces an electromagnetic field. For that reason, the power grid signal may be embedded also in some recorded signals when the device is connected to an electrical outlet. Alternatively, ENF contamination can be produced as a result of the presence of electromagnetic fields emanating from recorder power supply components such as transformers [6].

Assuming that the phase of the ENF varies within narrow limits, this work focuses on estimating abrupt changes in the ENF phase in order to determine whether or not the audio signal has been tampered with.

III. Proposed approach

The proposed system is depicted in the diagram of Figure 1. The first step is to preprocess the input signal and obtain the correspondent ENF tone. Later, as in a typical Phase-Locked Loop configuration, a Voltage-Controlled-Generator (VCO) produces a synthetic signal similar to the input. This voltage is then compared with the loop input. The differences between these two signals are used in order to correct the VCO-signal, making it closer to the ENF. In this sense, if the estimated ENF has strong phase variations, there will be a big difference between the loop input and the generated signals. This effect is reflected as a change in the VCO frequency value. Finally, the output of the system is the VCO frequency. The automatic decision about the audio authenticity is carried out comparing this frequency with an estimated threshold U. Each stage of the system is explained in the following subsections.

A. *Preprocessing*

Firstly, the signal was filtered with a low-pass filter of bandwidth 500 Hz and down-sampled to $f_s = 1000\ Hz$ in order to reduce the computational cost. This down-sampled signal is later filtered with a sharp band-pass filter (0.6 Hz bandwidth around the nominal ENF) so as to obtain the ENF.

The ENF has frequency and amplitude variations associated to the characteristics of the local power grid. These changes

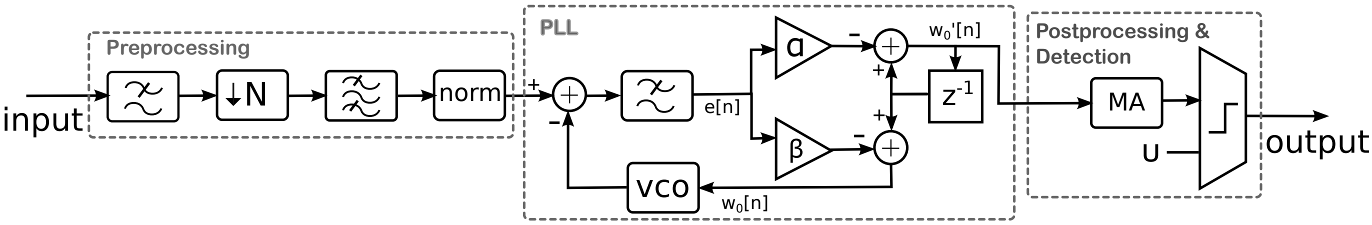

Fig. 1. Block diagram of the proposed system.

can introduce a certain degree of uncertainty, as the differences estimated by the PLL could be due to amplitude changes rather than phase changes. In order to have a correct performance of the PLL, a normalizing process is applied to the signal in order to have a constant amplitude. This normalization is carried out by dividing the signal by an estimation of its envelope (see Figure 2). The envelope of the signal is calculated using a full wave rectifier and a low-pass filter. A second order Chebyshev Filter is used due to its steep roll-off. The cut-off frequency of the LPF was set to 100 Hz.

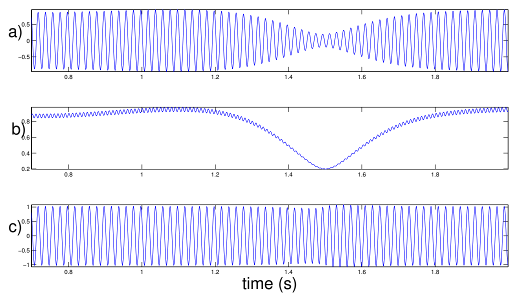

Fig. 2. Example of amplitude normalization using the envelope of the signal: a) signal b) envelope c) amplitude normalized signal.

B. The loop

The same low-pass filter is incorporated so as to smooth the error signal. In order to detect phase discontinuities of the ENF, the obtained signal after preprocessing is compared with a synthetic signal generated with a Voltage-Controlled-Generator (VCO). In addition, two pathways are included in the loop for controlling the update rate of the VCO parameters and the VCO output frequency (see Figure 1). Each block is detailed in the following.

1) VCO: In this case, the VCO was implemented with a typical sine-cosine generator [8], as shown in Figure 3. Recurrence equations of this system are:

$$y_1[n] = Cy_2[n-1] + Dy_1[n-1], \tag{1}$$

$$y_2[n] = By_1[n-1] + Ay_2[n-1]. \tag{2}$$

In order to determine its parameters such that the system behaves as a sine-cosine generator, unilateral Z-transforms

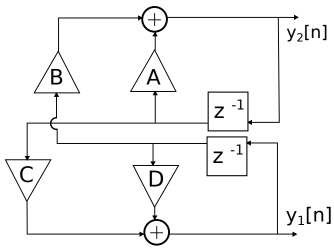

Fig. 3. Sine-cosine generator used to implement VCO.

(Z_u) were computed and the expressions of the system outputs are given as:

$$Y_{1_u}(z) = \frac{(Cy[0] - Ay_1[0])z^{-1} + y_1[0]}{1-(A+D)z^{-1} + (AD-BC)z^{-2}}, \tag{3}$$

$$Y_{2_u}(z) = \frac{(By_1[0]z^{-1} - +y_2[0] - Dy_2[0]z^{-1}}{1-(A+D)z^{-1} + (AD-BC)z^{-2}}. \tag{4}$$

The parameters were set as follows:

$$A = cos(\omega_0), \quad B = -sin(\omega_0), \quad C = sin(\omega_0), \quad D = cos(\omega_0),$$

and the initial conditions as: $y_1[0] = 1$ and $y_2[0] = 0$.

With this choice of parameters, the inverse transform of Equations (3) and (4) are, respectively:

$$y_1[n] = cos(\omega_0 n)u[n] \quad \text{and} \quad y_2[n] = sin(\omega_0 n)u[n],$$

where $u[n]$ is the Heaviside step.

2) The Pathways: The first pathway is used to update the VCO parameters (see $\omega_0[n]$ in Figure 1). This is carried out by using the difference between the VCO signal and the normalized ENF to correct the frequency of the VCO, which, in order to avoid sudden changes, has some inertia imposed by a first order low-pass filter. The updating expression of this pathway is given by:

$$\omega_0[n] = \omega_0'[n-1] - \alpha e[n], \tag{5}$$

where $\omega_0[n]$ is the VCO digital frequency at sample n, $\omega_0'[n]$ is an auxiliary frequency (explained next), α is a constant which value was set to $50/f_s$ (with f_s the sampling frequency), and $e[n]$ is the phase mismatch of the PLL. Equation (5) quickly updates the VCO parameters such that the VCO output frequency tracks the input signal frequency properly.

This updating process could generate strong oscillations around the input frequency and therefore loop instabilities.

In order to avoid this effect, a second pathway is introduced (see $\omega'_0[n]$ in Figure 1). This second updating expression is:

$$\omega'_0[n] = \omega'_0[n-1] - \beta e[n], \tag{6}$$

which is similar to Equation (5) but ω'_0 evolution is slower. Constant β was set to $\alpha/100$. The goal of this pathway is to smooth transitions in the output frequency of the VCO.

C. Postprocessing

In order to determine the audio authenticity automatically, the VCO output frequency is compared with a threshold, as explained in Section III-D. To ensure that this threshold performs properly, a postprocessing stage was included. As shown in Figure 4, the ENF presents variations that could be considered as phase discontinuity, a false detection. This kind of variation can appear when the ENF component is interfered with another signal, such as with a click. This produces an abrupt variation of the ENF which the algorithm detects as a false positive. Those variations present a typical pattern: they have positive and negative components with respect to the mean value. Because of interference, the PLL looses the phase of the signal, measuring an offset. As the actual phase of the signal did not change, in order to lock back into it, a phase difference with the opposite sign appears later. This does not happen in a true positive case as the PLL will make a single change in order to be locked to the new phase.

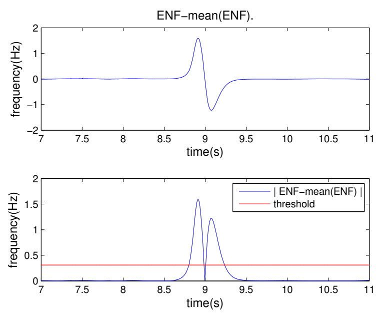

Fig. 4. PLL output in case of false a positive.

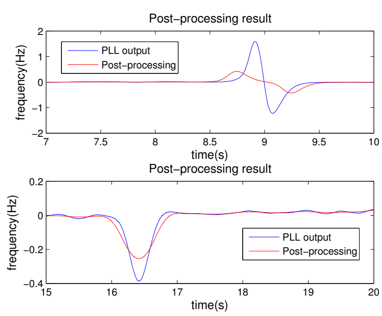

Fig. 5. Result with postprocessing.

To postprocess the signal, a moving average (MA) filter was employed with a window length corresponding to 500ms. Therefore, if the signal has positive and negative variations, its amplitude is decreased by the MA filter more than if the signal has only positive or only negative variations. Figure 5 shows the filter result for the two cases. In the upper plot, there is a false negative while, in the bottom plot, a true positive is shown. The results comparing the system performance with and without postprocessing are presented in Section IV.

D. Automatic decision

In order to discriminate between edited and original signals, it is necessary to characterize abrupt changes in the ENF. As a final step of the proposed system, the VCO output frequency is compared to a threshold in order to determine automatically if the audio has been edited or not. This threshold may be computed taking into account the Equal Error Ratio (EER), as in previous works [3], [9].

The algorithm was tested with two databases, CARIOCA [3] and AHUMADA [7]. Each database has 200 audio signals, half of those are original and the other half are edited. In Figures 6 and 7, the Detection Error Tradeoff (DET) curves for each database are shown.

As shown in the DET curves, two threshold values were selected in order to compare system results with and without postprocessing. In case of CARIOCA database, the postprocessing improves the EER from 5% to 4%; conversely, in AHUMADA database, there is no significant difference. The experimental results are presented in Section IV.

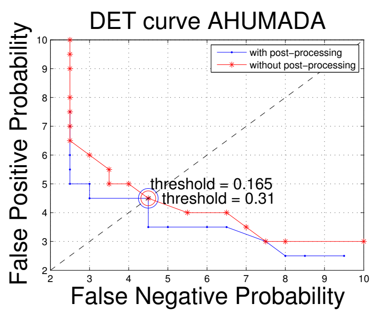

Fig. 6. Threshold value selection using EER (in %) in AHUMADA database.

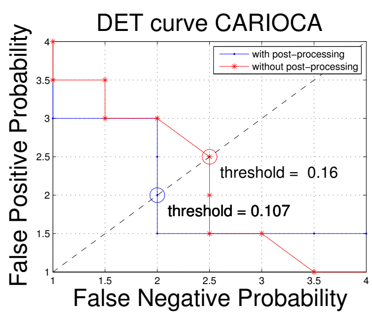

Fig. 7. Threshold value selection using EER (in %) in CARIOCA database.

IV. Experiments and Results

As before mentioned, the system was trained with two databases (AHUMADA and CARIOCA) and an EER threshold was established for each one. To test the generalization ability of the system, each threshold was considered for testing the other database, in a Cross Validation (CV) scheme. The results are presented in Table I, comparing the system performance with and without postprocessing. An important issue to highlight is that the FP and FN probabilities presented in Table I are different from those shown in Figures 6 and 7. In that case, each FN and FP probability were computed for each value of threshold, which was varying to form the DET curve (the EER being one point of the curve, useful to have a glimpse of the system performance but not necessarily the operating point to be chosen in every implementation).

As shown in Table I, the algorithm performance achieves 96% of correct detections in CARIOCA database and 89% in AHUMADA. This difference could be explained taking into account that AHUMADA database contains about 10 (5% of the *corpus*) distorted signals [9].

TABLE I
ALGORITHM ERROR PROBABILITY IN A CV SCHEME

	Carioca train Ahumada test		Ahumada train Carioca test	
	w/o pos	w/ pos	w/o pos	w/ pos
FP (%)	8.5	8.5	2.0	0.5
FN (%)	2.5	2.5	2.5	3.5
Error (%)	11.0	11.0	4.5	4.0

It would be interesting to know the robustness of this algorithm in the presence of noise and nonlinear distortion. For this purpose two extra experiments were carried out. Firstly, the algorithm was tested with clipped signals in 0.5% of samples and secondly with the signals contaminated with white noise (SNR of 10 dB, assuming the original recordings as clean signals). In this case the point of operation is the EER in each database. The results are presented in Table II. As shown in this Table, when compared to previous results, the EER has improved by 1.5% for the *clean* Ahumada corpus (from 6% in [3] to 4.5% here) and by 2% for the *clean* Carioca Corpus (from 4% in [9] to 2% here).

TABLE II
ALGORITHM ERROR PROBABILITY: EFFECT OF CLIPPING (**0.5%**) AND WHITE GAUSSIAN NOISE (**SNR 10**dB)

	Ahumada			Carioca		
	ori	clipp	noise	ori	clipp	noise
FP (%)	4.5	27.5	49	2	33.5	50
FN (%)	4.5	2.5	0.5	2	0.5	0
Error (%)	9	30	49.5	4	34	50

V. CONCLUSION

A novel PLL based method for determining audio authenticity was proposed in this work; several experiments were conducted in order to evaluating its generalization ability. Two different databases containing both original and edited signals were used to show that the method is able to pinpoint ENF discontinuities with good accuracy, achieving 2% EER in CARIOCA database and 4.5% EER for the AHUMADA database. This performance is in line with those attained in previous works [3], [9], with a mean improvement of 1.75% of the EER.

A cross validation scheme was carried out, obtaining a 96% of correct detections when training with AHUMADA and testing with CARIOCA. When training with CARIOCA and testing with AHUMADA, a 89% of correct detections is performed. These experiments yielded promising results that show the advantages of the proposed method.

Additional tests were conducted to estimate the system performance in presence of noise and clipping: the signals were distorted with amplitude clipping (0.5%) and corrupted with additive white gaussian noise (SNR 10dB). The results obtained show that these distortions severely affect the performance of the proposed method. Another important issue to take into account is related to pathological cases (e.g. when the edition length is an integer number of ENF cycles); in those cases, it is not possible to employ any ENF-based audio authenticity method.

As future work, a larger database could be formed with each signal containing the time information of each edition, helpful to determine if the algorithm recognizes a real discontinuity or if it detects spurious ENF variations. Moreover, a combined approach—PLL, phase, and instantaneous frequency—is a natural candidate for future research.

ACKNOWLEDGMENT

Authors thank Mr. L. Hirsch and Prof. M. Moeneclaey for recommending the use of PLL in this application.

REFERENCES

[1] R. C. Maher, "Audio forensic examination: Authenticity, enhancement, and interpratation," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 84-94, 2009.

[2] R. W. Sanders, "Digital authenticity using the electric network frequency," In Proc. of the *33rd AES International Conference: Audio Forensics, Theory and Practice*, Denver, USA, June 2008, pp. 1–11.

[3] D. P. N. Rodríguez, J. A. Apolinário Jr., and L. W. P. Biscainho, "Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 534–543, 2010.

[4] D. P. Nicolalde and J. A. Apolinário Jr., "Evaluating digital audio authenticity with spectral distances and ENF phase change," In Proc. of the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, April 2009, pp. 1417-1420.

[5] A. J. Cooper, "An automated approach to the Electric Network Frequency (ENF) criterion: Theory and practice," *International Journal of Speech, Language and the Law*, vol. 16, no. 2, pp. 193–218, 2009.

[6] E. B. Brixen, "ENF quantification of the magnetic field," In Proc. of the *33rd AES International Conference: Audio Forensics, Theory and Practice*, Denver, USA, June 2008, pp. 1–6.

[7] J. Ortega-García, J. Gonzalez-Rodríguez, and V. Marrero-Aguiar, "AHUMADA: A large speech corpus in Spanish for speaker characterization and identification," *Elsevier Science B.V., Speech Communication*, vol. 31, pp. 255–264, 2000.

[8] J. W. Gordon and J. O. Smith, "A Sine Generation Algorithm for VLSI Applications," In Proc. of the *International Computer Conference (Computer Music Association)*, vol. 1985, Vancouver, Canada, 1985, pp. 165–168.

[9] P. A. A. Esquef, J. A. Apolinário Jr., and L. W. P. Biscainho, "Edit Detection in Speech Recordings via Instantaneous Electric Network Frequency Variations," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2314–2326, 2014.