Electric Network Frequency Estimation Based on Linear Canonical Transform for Audio Signal Authentication

Wei Zhong and Xingang You
Beijing Institute of Electronics Technology and Application
Beijing, P. R. China, 100091
Email: zww110221@163.com
and youxg@dlut.edu.cn

Xiangwei Kong and Bo Wang
School of Information and Communication Engineering
Dalian University of Technology,
Liaoning, P.R. China, 116024
Email: kongxw@dlut.edu.cn and bowang@dlut.edu.cn

***Abstract*—As electric network frequency is sometimes embedded in audio signals when the recording is carried out with the equipment connected to an electrical outlet, electric network frequency estimation is an important task in audio authenticity. After the theoretical analysis, a novel electric network frequency estimation method based on linear canonical transform is proposed from anti-multipath interference point of view. The experimental results demonstrate that this model performs well with high precision in complex noisy environment.**

I. Introduction

As one of the most important multimedia in this era, digital audio plays a significant role in the modern lives. It is very common for the individuals to use recorders, cameras and smartphones to collect digital audios. And also, the large-scale social media network makes it possible to spread the digital audio rapidly. The convenience of acquiring and propagating the digial audios also brings a potential risk for the content security of digital multimedia. The digital audio could be easily modified and tampered via the widely spread and sophisticated tools, which means, even an amateur could create a forgery audio, and sequentially spread it via the Internet and social media applications, Facebook and Twitter for instance. In some scenarios, the forgered audios would endanger the impartiality of judiciary.

Based on this consideration, audio forensics has found its application in the law court [1] since 1960s especially when the Water Gate was exposed [2]. Some literatures have focused on the digital audio forensics, especially in digital audio source identification and digital audio forgery detection. As a unique attribution of digital multimedia currently embedded into the audio signals, electric network frequency (ENF) is usually connect to different electrical equipments when the recording is carried out. Thus, the estimation of electric network frequency is widely used as an important tools for audio authenticity. In recent years, several approaches have been exploited in previous works. Grigoras [3] first found the components of electric network frequencies in an audio recording. By estimating the electric network frequencies, several features were proposed and compared with the power network frequency database. A high correlation was determined to expose the recording data. Furthermore, Brixen [4] verified the correlations by numerous experiments, and applied the techniques to the digital audio forgery detection. Besides, they found that the ENF also provided an opportunity to solve the problem of forgery detection in the same sample ratio.

Nicolalde $et\ al.$ [5, 6] used high precision phase estimation method to determine the discontinuity of the audio recording and locate the audio manipulation points. M.Huijbregtse [7] $et\ al$ studied the relationship between the frequency curve of a power network and the audio recording time. The adaptation of power supply characteristics to the audio forensic has made fast and accurate frequency estimation becoming popular study subjects.

All of the traditional algorithms, LMS, Least Square Algorithm, Kalman FItering for instance, have the disadvantages of long delay for frequency estimation, due to the fact that extra filters are used in the algorithms. PHD and MUSIC algorithms have computing limitations due to their high complexity. Hing Cheung So $et\ al$ [8] proposed a fast and accurate frequency estimation method in the sub-space domain, while Hyeon-Jin Jeon $et\ al$ [9] proposed iterative frequency estimation based on MVDR spectrum.

The power supply networks in Europe and U.S.A. are generally of less noisy. The published frequency estimation algorithms are effective for these countries. However they are less effective for the power distribution networks in China due to noisier environment. More robust and anti-noise algorithms must be derived to estimate the frequency. Linear canonical transform was first proposed by Moshinsky $et\ al$ [10] in 1970s. Due to its flexible rotation feature, it has been widely used in the area of signal processing. A novel electric network frequency estimation method based on linear canonical transform is proposed in this paper from anti-multipath interference point of view. According to the theoretical analysis and experimental result, this model outperforms existing algorithm with high precision in complex noisy environment.

The paper is organized as follows. Section II provides frequency estimation for complex noisy electrical power grid.

ISBN 978-0-9928626-7-1 © EURASIP 2017

Section III discusses the calculations for coefficients a, b, c, d of applying linear canonical transform for a narrow band, amplitude-variant signal. Section IV will give the experimental results, which applies LCT method of frequency estimation to various complex noisy environments, are demonstrated in Section V. And also the comparison with baselines, the traditional DFT based algorithm, is made in this Section. Finally, the conclusion is derived in Section V.

II. Frequency Estimation for Complex Noisy Electrical Power Grid

This section starts with the analysis on the frequency estimation from ideal electrical power signal and the complex noisy power network.

For complex noisy electrical power grid, not only the white noise, but also the multi-path noise exist. Normal frequency estimation algorithms will become less accurate if this complex noise occurs.

Assuming that the electrical signal could be modeled as the following formula:

$$s\left(t\right) = \lambda_0 e^{j\omega_c t} \tag{1}$$

Take the derivative on formula (1):

$$s'\left(t\right) = j\lambda_0 e^{j\omega_c t}\omega_c \tag{2}$$

And the following formulas hold:

$$\frac{1}{T}\int\limits_0^T s^*\left(t\right)s\left(t\right)dt = \left|\lambda_0\right|^2 \tag{3}$$

$$\frac{1}{T}\int\limits_0^T s^*\left(t\right)s'\left(t\right)dt = j\omega_c\left|\lambda_0\right|^2 \tag{4}$$

$s^*(t)$ is the complex conjugate of $s(t)$, the estimation of the working frequency then may be obtained as:

$$\hat{\omega}_c = im\left(\frac{\int_0^T s^*(t)s'(t)dt}{\int_0^T s^*(t)s(t)dt}\right) \tag{5}$$

Bykhovsky and Cohen [11] proposed a model of electric network frequency in 2013, as Eq.(6) illustrated below. b_0 denotes the DC component in the channel, b_{cm} denotes different amplitudes of harmonics, M denotes the number of harmonics and $n(t)$ is the noise of the electric network. Giving consideration to both side of the harmonics components and noise, this model is widely accepted in the Europe and the United States. In spite of this, we do believe the model is inadequacy for the electric network in China, because this model ignores the randomness of delays and amplitudes. Considering the complex and noisy electric network in China, the electrical signal becomes a narrow band, amplitude-variant signal which can be expressed as follows:

$$r(t) = \sum_{i=1}^{P}\lambda_i\exp(j\eta_i t)\exp[j\varphi_i(t)] + n(t) \tag{7}$$

where P is the number of multipath and $n(t)$ is the narrow band Gaussian noise.

$$\varphi_i(t) = \theta_i + \omega_c t + \frac{\phi_i}{2}t^2 \tag{8}$$

Substitute Eq. (7) into Eq.(3) to (5), we can the estimated $\hat{\omega}_c$ carries a large estimation error as shown in Eq.(9).

It is natural therefore that the original signal may be transformed into some other domain before the estimation is done. Eq. (7) shows the signal is band-limited and composite signal in time domain. Applying linear canonical transform with appropriate coefficients a, b, c, d, this signal can be transformed into a frequency spectrum with a sharply pulse.

$$r(t) = b_0 + \sum_{m=1}^{M} b_{cm}\cos(\omega_0 mt) + \sum_{m=1}^{M} b_{cm}sin(\omega_0 mt) + n(t) \tag{6}$$

$$\hat{\omega}_c = im\left(\frac{\int_0^T \sum\limits_{i=1}^{P}\sum\limits_{k=1}^{P}\lambda_i\lambda_k\exp\left(j[(\theta_k-\theta_i)+j(\eta_k-\eta_i)t+\frac{\phi_k-\phi_i}{2}t^2]\right)(j\omega_c+jc_k+\phi_k t)\,dt}{\int_0^T \sum\limits_{i=1}^{P}\sum\limits_{k=1}^{P} b_i b_k\exp\left(j[(\theta_k-\theta_i)+j(\eta_k-\eta_i)t+\frac{\phi_k-\phi_i}{2}t^2]\right)dt}\right) \tag{9}$$

$$R_{a,b,c,d}(u) = \begin{cases} \sqrt{\frac{1}{j2\pi b}}\exp\left(\frac{jd}{2b}u^2\right)\int\limits_{-\infty}^{+\infty}\exp\left(\frac{ja}{2b}t^2-\frac{j}{b}ut\right)r(t)dt & ad-bc=1 \\ \sqrt{d}\exp(\frac{jcd}{2}u^2)f(du) & b=0 \end{cases} \tag{10}$$

$$R_{a,b,c,d}(u) = \sum_{i=1}^{P}\sqrt{\frac{1}{a+\phi_i b}}\lambda_i\exp(jA_i)\exp(jB_i u)\exp(jC_i u^2) + N(u) \tag{11}$$

$$A_i = \theta_i - \frac{(\omega_c+\eta_i)^2 b}{(a+\phi_i b)}, B_i = \frac{(\omega_c+\eta_i)}{(a+\phi_i b)}, C_i = \left(\frac{d-a-\phi_i b}{2b(a+\phi_i b)}\right) \tag{12}$$

ISBN 978-0-9928626-7-1 © EURASIP 2017

The linear canonical transform (LCT) takes the form Eq.(10) as follows.

Applying LCT to Eq.(7), we subsequently obtain Eq.(11). $N(u)$ is the LCT for $n(t)$, A_i, C_i, B_i are defined in Eq.(12).

Setting C_i=0,we can obtain a, b, c, d by $d - a - \phi_0 b = 0$. And ϕ_0 is the 2nd-order coefficient corresponding to the max distance. Therefor, Eq.(13) could be derived from Eq.(11) as follows.

By applying a narrow band filter with the central frequency $u_0 = (\omega_c + \eta_0), /d$, we obtain:

$$R_0(u) \approx \sqrt{\frac{1}{d}}\lambda_0 \exp(jA_0) \exp(j\frac{\omega_c + \eta_0}{d}u) \tag{14}$$

which shows the signal becomes a single tone, as a predictable result of the narrow band filtering. In other words, the energy aggregation is realized by this way. Also, we could get Eq. (15) and (16) according to the characteristics of LCT.

$$\frac{1}{T}\int_0^T R^*{}_0(u)R_0(u)du = \frac{|\lambda_0|^2}{d} \tag{15}$$

$$\frac{1}{T}\int_0^T R^*{}_0(u)R_0'(u)du = j\frac{|\lambda_0|^2}{d^2}(\omega_c + \eta_0) \tag{16}$$

Finally, we obtain the estimated frequency using LCT under a complex noisy environment, as Eq.(17) demonstrated.

$$\hat{\omega}_c = im\left(\frac{\int_0^T R_0 * (u)R_0'(u)du}{\int_0^T R_0 * (u)R_0(u)du}\right)d - \eta_0 \tag{17}$$

III. Calculations of Coefficients

Considering Eq.(7) and Eq.(10), the max energy distance in LCT has a sharp pulse when $a_1/b_1 = \phi_0$. Eq.(14) also has a high peak after the Fourier Transform. Therefore the following equation holds:

$$\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{18}$$

So that:

$$\begin{cases} ad - bc = 1 \\ \frac{c}{b} = \phi_0 \\ a + b\phi_0 = d \end{cases} \tag{19}$$

And a, b, c, d are calculated from Eq.(19).

TABLE I
Comparison of Estimation Algorithms for $f_c = 50.58Hz$

	DFT	Derivative	LCT
L = 128	54.6875	50.0759	50.6162
L = 256	54.6875	50.1256	50.6161
L = 512	52.7344	50.3991	50.5988
L = 1024	51.7578	50.4686	50.5899
L = 2048	51.2695	50.5759	50.5861
AVG.	53.0273	50.3290	50.6014

TABLE II
Comparison of Estimation Algorithms for $f_c = 50.58Hz$ With $5dB$ Gaussian White Noises

	DFT	Derivative	LCT
L = 128	54.6875	50.0741	50.6137
L = 256	54.6875	50.1214	50.6154
L = 512	52.7344	50.3464	50.5987
L = 1024	51.7578	50.4807	50.5904
L = 2048	51.2695	50.5730	50.5859
AVG.	53.0273	50.3191	50.6008

IV. Experiment and Results

To evaluate the performance of the proposed frequency estimation method using LCT, simulations are carried out under various complex and noisy environments in our experiments, with the traditional DFT and Derivation algorithms as the baselines for comparison.

Let k_{peak} is the max integer of N-point DFT $|R(k)|$ of $r(t)$, so that:

$$f_{DFT} = k_{peak}\frac{f_s}{N_{DFT}} \tag{20}$$

Several digital audio pieces, which are collected from young male and female voices, are used in our experiments as test samples. All of the samples are recorded with $8kHz$ sampling frequency and with 8 bits resolution. To simulate the various complex and noisy environments, the original voices are added with different noises, the gaussian white noise and pink noise for instance. The guassian white noises are supplied by RSRE (The Royal Signals and Radar Establishment) speech research center from the United Kingdom.

A. Single Frequency Estimation

A sinusoidal signal f_c=50.58Hz is generated with sampling frequency 1000Hz. DFT, Derivation and LCT based estimation algorithms are carried on the sinusoidal signal. Table 1 gives the estimation results of the three methods for different sample data length L.

$$R_{a,b,c,d}(u) = \sqrt{\frac{1}{d}}\lambda_0 \exp(jA_0) \exp(j\frac{\omega_c + \eta_0}{d}u) + R_{a,b,c,d,i\neq \max I}(u) + N(u) \tag{13}$$

TABLE III
COMPARISON OF ESTIMATION ALGORITHMS FOR $f_c = 50.58Hz$ WITH $5dB$ PINK NOISES

	DFT	Derivative	LCT
L = 128	54.5277	50.0047	50.6175
L = 256	54.6875	50.1159	50.6140
L = 512	52.7344	50.3629	50.5983
L = 1024	51.7578	50.4819	50.5929
L = 2048	51.2695	50.0599	50.5869
AVG.	52.9954	50.2051	50.6019

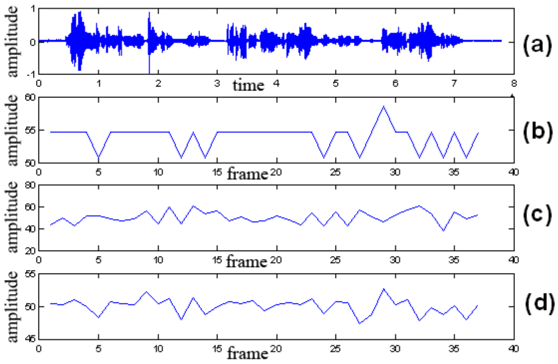

Fig. 1. Noise Free Performance Comparison: (a) Testing waveform, (b) DFT method, (c) Derivative Method, (d) LCT Method.

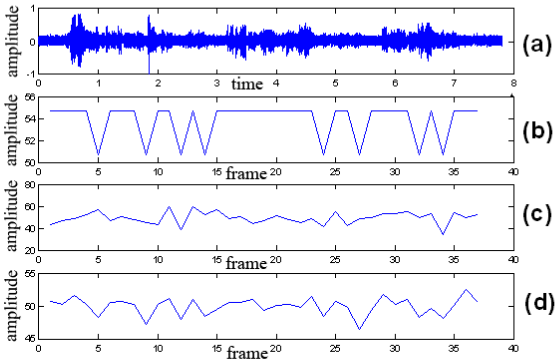

Fig. 2. Noise ($5dB$ Gaussian) Performance Comparison: (a) Testing waveform, (b) DFT method, (c) Derivative Method, (d) LCT Method.

As the Table 1 demonstrates, decent average frequency of $53.0273Hz$ and $50.3290Hz$ are achieved of DFT and Derivation, while the proposed LCT algorithm obtains a more accurate frequency of $50.6014Hz$, for the ideal and noiseless environment. The table also shows that, with longer sample data length, the estimation results of LCT improve. In the case of L=128, the estimated frequency is $50.6162Hz$ while in the case of L=2018, the result is $50.5861Hz$.

Similar results could be found in Table 2 and 3, in the $5dB$ gaussian white noise and $5dB$ pink noise scenarios. We both find that the LCT makes the best estimation results, the

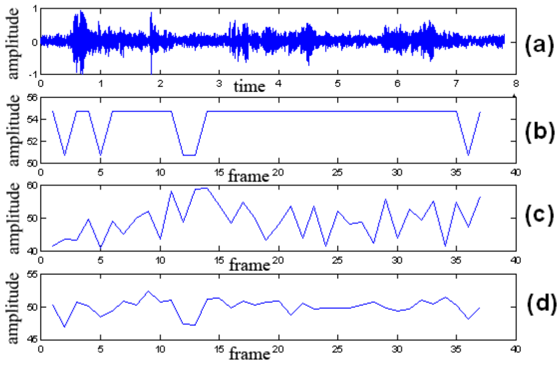

Fig. 3. Noise (Factory noise) Performance Comparison: (a) Testing waveform, (b) DFT method, (c) Derivative Method, (d) LCT Method.

Derivation and DFT are worse than that of LCT. In the case of L=2048, LCT gets an estimation of $50.5859Hz$ with the gaussian white noise, and $50.5869Hz$ with the pink noise, while in the same condition, the results of DFT algorithm are both $51.2695Hz$ and the results of Derivation are $50.5730Hz$ and $50.0599Hz$ seperately.

B. Power Line Frequency Estimation in Speech

The performance of estimation of frequency in power line are also compared in our experiments, where L is fixed to 256 points. Fig. 1 illustrates the results of noise free environment. Fig. 2 and Fig. 3 show the results in white noise and factory noise environments, respectively. Also without doubt, LCT algorithm proposed in this paper outperforms the other two baselines.

V. CONCLUSION

Frequency estimation of power line plays key role for the audio signal forensics. A robust and accurate algorithm based on the linear canonical transform is proposed in this paper. The experimental results show that the novel algorithm is superior to the conventional DFT and Derivative Transform based methods. And also, the simulations indicate that the proposed LCT algorithm shows high robustness in the complex and noisy environments, the power line for instance.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No.61502076, and also the Scientific Research Project of Liaoning Provincial Education Department, No.L2015114.

REFERENCES

[1] A. J. Cooper, “Detection of Copies of Digital Audio Recordings for Forensic Purposes”, *Milton Keynes, the Open University*, March, 2006.

[2] Advisory Panel on White House Tapes. The Executive Office Building Tape of June 20, 1972: Report on a technical investigation.United States District Court for the District of Columbia, May 31,1974. Available:http://www.aes.org/aeshc/docs/forensic audio/watergate. tapes.Report. Pdf . 1974.

[3] C. Grigoras. “Application of ENF analysis method in authentication of digital audio and video recordings”, in *Proc. Audio Engineering Society 123rd Conv.*, New York, 2007.

[4] E. B. Brixen, “ENF;Quantification of the magnetic field,” in *Proc. 33rd Conf. Audio Forensics-Theory and Practice*, Denver, CO, June, 2008.

[5] D. P. Nicolalde, J. A. Apolinrio, and L. W. P. Biscainho, “Audio authenticity: Detecting ENF discontinuity with high precision phase analysis,” *IEEE Trans. on Inform. Forensics and Security*, vol. 5, no.3, pp. 534-543, June 2010.

[6] D. P. Nicolalde and J. A. Apolinario, “Evaluating digital audio authenticity with spectral distances and ENF phase change”, in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, May 2009.

[7] B. Dam, J. M. Huijbregtse, and F. C. Klaassen, et al, “Origin of high critical currents in superconducting thin films“, *Nature*,vol. 399, no. 6735, pp. 439-442, 1999.

[8] H. C. So, F.K.W. Chan, and Weize Sun, “Subspace Approach for Fast and Accurate Single-Tone Frequency Estimation”, *IEEE Trans. on Signal Processing*, vol. 59, no. 2, pp. 827-831, Nov. 2010.

[9] H-J Jeon, and T-G Chang, “Iterative Frequency Estimation Based on MVDR Spectrum”, *IEEE Trans on Power Delivery*, vol. 25, no. 2, pp. 21-630, March 2010.

[10] M. Moshinsky and C. Quesne, “Linear cannonical transformations and their unitary representations”. *J. Math. Phys.*, vol. 12, no. 8, pp. 1772-1783, 1971.

[11] D. Bykhovsky, and A. Cohen, “Electrical network frequency (ENF) maximum-likelihood estimation via a multitone harmonic model”, *IEEE Trans. on Inform. Forensics and Security*, vol. 8, no. 5, pp. 744-753, March 2013.