EXPLORING THE USE OF ENF FOR MULTIMEDIA SYNCHRONIZATION

Hui Su, Adi Hajj-Ahmad, Min Wu, and Douglas W. Oard
{hsu, adiha, minwu, oard}@umd.edu
University of Maryland, College Park

ABSTRACT

The electric network frequency (ENF) signal can be captured in multimedia recordings due to electromagnetic influences from the power grid at the time of recording. Recent work has exploited the ENF signals for forensic applications, such as authenticating and detecting forgery of ENF-containing multimedia signals, and inferring their time and location of creation. In this paper, we explore a new potential of ENF signals for automatic synchronization of audio and video. The ENF signal as a time-varying random process can be used as a timing fingerprint of multimedia signals. Synchronization of audio and video recordings can be achieved by aligning their embedded ENF signals. We demonstrate the proposed scheme with two applications: multi-view video synchronization and synchronization of historical audio recordings. The experimental results show the ENF based synchronization approach is effective, and has the potential to solve problems that are intractable by other existing methods.

Index Terms— ENF, synchronization, audio, video, historical recordings

1. INTRODUCTION

The analysis of electric network frequency (ENF) signals has emerged in recent years as an important technique for digital multimedia forensics. ENF is the supply frequency of power distribution networks in a power grid. The nominal value of the ENF is usually 60Hz (in North America) or 50Hz (in most other parts of the world). The instantaneous value of ENF fluctuates slightly around its nominal value due to load variations and the control mechanisms of the power grids. The main trends in the fluctuations of the ENF have been shown to be very similar within the same power grid. The changing values of the ENF over time are regarded as the ENF signal. The ENF signal can be extracted from power signals measured from a power outlet using a step-down transformer and a simple voltage divider circuit.

Multimedia recordings created using devices plugged into the power mains or located near power sources can pick up ENF signals in audio due to electromagnetic interference or acoustic vibrations [1]; and in video due to imperceptible flickering in indoor lighting [2]. The ENF signal extracted from audio or video recordings has been shown to exhibit a high correlation with the ENF extracted from the power mains measurements at the corresponding time. Several forensic applications have been proposed based on the analysis of the ENF signal. For example, ENF signals have been successfully used as a natural time stamp to authenticate audio recordings [3, 1, 4]. By examining the phase continuity of the ENF signal, one can detect the region of tampering [5]. Some recent work shows that the ENF signal can also reveal information about the locations and regions in which certain recordings are made [6, 7, 8].

In this paper, we explore the potential of the ENF signal from a new perspective and use it for synchronization of multimedia signals, i.e. to temporally align audio and video recordings. Synchronization is a fundamental problem for applications dealing with multiple pieces of multimedia signals such as view synthesis and A/V experience reconstruction [9]. Existing approaches to multimedia signal synchronization, which generally extract and match audio/visual features, may not always work well. For example, it is difficult to synchronize video sequences using visual features when they do not share sufficient common view of the scene; similar limitations apply to alignment of audio recordings that have no common acoustic or speech events.

The ENF signal is a continuous random process over time. Multimedia recordings can therefore be synchronized by aligning their embedded ENF signals. As this method does not rely on the audio or visual information of the multimedia signals, it is complementary to the conventional synchronization approaches, and it may help to solve problems that are otherwise intractable. The rest of the paper is organized as follows. Section 2 describes the basic methodology of the proposed idea. Then we demonstrate this approach with two applications. Section 3 shows examples of multi-view video synchronization using the ENF signal extracted from soundtracks. In Section 4, the proposed method is applied to synchronize some audio recordings of historical importance. Section 5 concludes the paper.

2. METHODOLOGY

2.1. Extraction of the ENF Signal

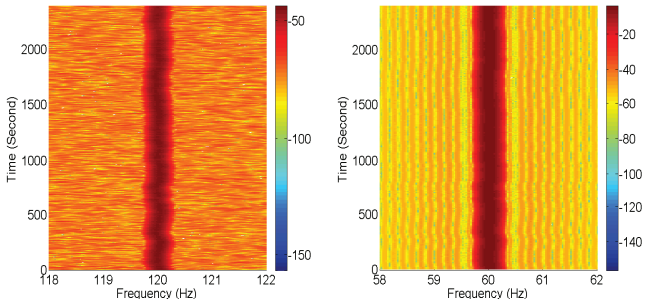
The ENF signal embedded in multimedia recordings is usually present around its nominal value and the higher order har-

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

monics. In Fig.1 (a) and (b) of the spectrograms of an audio signal and the power mains measurement signal recorded at the same time, we observe a strip of time-varying energy at 120 Hz and 60 Hz, respectively, which correspond to the ENF signals in these recordings. We can extract the ENF signal by estimating the instantaneous peak frequency among a small range ($\pm\Delta f$) around the ENF nominal value and harmonics. Comparisons of various frequency estimation approaches for ENF were carried out in [8, 10]. The weighted energy method [2] is adopted here for its robustness and low complexity. The recording signals are divided into frames of certain length (e.g, 8 seconds), and FFT is calculated for every frame. The ENF signal is then estimated by:

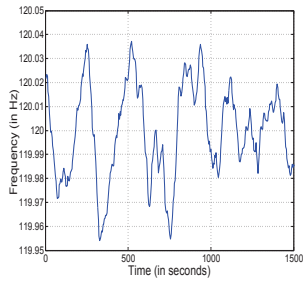
$$F(n) = \frac{\sum_{l=L_1}^{L_2} f(n,l)|s(n,l)|}{\sum_{l=L_1}^{L_2} |s(n,l)|}, \qquad (1)$$

where f_s and N_{FFT} are the sampling frequency of the signal and the number of FFT points, respectively; $L_1 = \frac{(f_{ENF}-\Delta f)N_{FFT}}{f_s}$ and $L_2 = \frac{(f_{ENF}+\Delta f)N_{FFT}}{f_s}$; $f(n,l)$ and $s(n,l)$ are the frequency and energy in the l^{th} frequency bin of the n^{th} time frame, respectively. Fig.1 (c) and (d) show the ENF signals estimated from the audio recording and the concurrent power signal, and the two have very similar fluctuation trends.

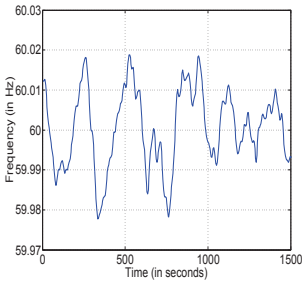


(a) Spectrogram of an audio signal at 2^{nd} harmonic (120 Hz).

(b) Spectrogram of the corresponding power signal at 60 Hz.



(c) ENF signal estimated from the audio signal.



(d) ENF signal estimated from the power signal.

Fig. 1. Spectrograms and ENF estimates from audio and power signals recorded at the same time.

2.2. Synchronization using ENF

The value of the ENF fluctuates around its nominal value due to varying supply and loads over the power grids. The major trends of these fluctuations are consistent at all locations across the same grid. Previous work has exploited the property of the ENF traces embedded in multimedia recordings for digital forensic purposes. In this paper, we explore the utilization of the ENF signals in multimedia recordings from a new perspective. In viewing the ENF signal as a continuous-time random process, its realization in each recording may serve as a timing fingerprint. Synchronization of audio and video recordings can therefore be performed by matching and aligning their embedded ENF signals. This is a very different approach to tackling the audio/video synchronization problem from existing work, and has several advantages over conventional methods. The ENF based method does not reply on having common audio and visual contents between the multiple recordings to be synchronized. Taking video synchronization for example, the conventional approaches based on visual cues do not work well in situations where there are arbitrary camera motions or the view overlap is insufficient, while the ENF based method is not affected by these adverse conditions. Additionally, extracting and aligning ENF signals may be more effective computationally than the approaches that rely on computer vision and/or extensive learning, and thus more (or longer) recordings could be efficiently processed. It can also be easily generalized to synchronize multiple pieces of recordings.

There are several requirements for the ENF based synchronization approach to work. The ENF traces in the audio and video recordings must be strong enough so that reliable ENF signals can be estimated. The temporal overlap between recordings to be synchronized should be sufficiently large to ensure accurate alignment of the ENF signals. These requirements may not be always satisfied. In our experiments, we find the proposed method can work well in diverse settings. In the following sections, we demonstrate the performance of the ENF based synchronization with audio-video files and historical audio recordings.

3. ENF FOR VIDEO SYNCHRONIZATION

In this section, we discuss in details how the ENF traces embedded in video soundtracks can be used for video synchronization. After taking the soundtracks from two video recordings to be synchronized, we first divide each soundtrack into overlapping frames of length L_{frame} seconds. The overlap between adjacent frames is denoted as $L_{overlap}$ in seconds. So the shift from one frame to the next is $L_{shift} = L_{frame} - L_{overlap}$. For every frame, we estimate the dominant frequency around the nominal value of the ENF. The values of the estimated frequency are concatenated together to form the ENF signal of each soundtrack. The normalized cross corre-

Table 1. Synchronization accuracy with fixed L_{shift} of 1 second and varying L_{frame}

L_{frame} $(sec.)$	8	16	24	32
$RMSE$ $(sec.)$	0.79	0.33	0.32	0.33
MAE $(sec.)$	0.46	0.27	0.27	0.27

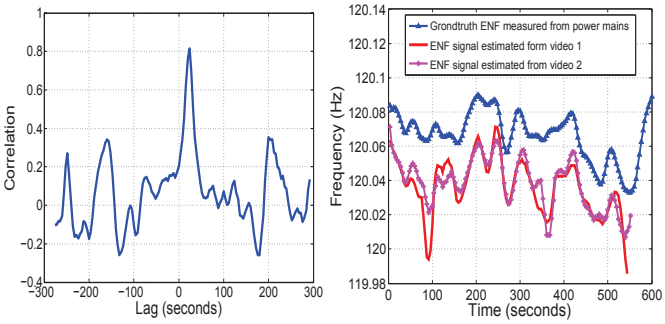
Table 2. Synchronization accuracy with fixed L_{frame} of 16 seconds and varying L_{shift}

L_{shift} $(sec.)$	1	0.5	0.3	0.1
$RMSE$ $(sec.)$	0.33	0.21	0.17	0.15
MAE $(sec.)$	0.272	0.166	0.136	0.117

lation coefficients are calculated with different lags between the ENF signals. The lag corresponding to the maximum correlation coefficients is identified as the temporal shift between the two videos.

The accuracy of synchronization is important for many applications involving multiple videos. Experiments are conducted to examine the accuracy of the proposed method. We take multiple video clips simultaneously with two different cameras at different locations, including offices, hallways, recreation centers and lobbies. These videos are divided into segments of 10 minutes long and each segment is treated as a test sample. The soundtracks of the segments are analyzed and the ENF signals are extracted from them for synchronization. The ground truth of the lag between the recordings was obtained by manually comparing the video frames, and used to measure the synchronization accuracy in terms of mean absolute error (MAE) and root mean square error (RMSE) under different settings of L_{frame} and L_{shift}. The experimental results are listed in Table 1 and 2. We first fix L_{shift} as 1 second and test different values of frame length L_{frame}. The alignment accuracy becomes better when L_{frame} is increased, and becomes saturated at the frame length of 16 seconds or longer. Next, L_{frame} is fixed as 16 seconds, and L_{shift} is varied from 1 second down to 0.1 second. The synchronization accuracy improves as we use a smaller L_{shift}. With $L_{frame} = 16, L_{shift} = 0.1$, the MAE is about 0.12 second, equivalent to 3.6 frames for videos of 30 frames / second.

Fig. 2 shows an example of video synchronization using the proposed approach. We use two cameras to video tape a racket ball court from two different angles. Fig. 2 (a) is the correlation coefficients between the ENF signals extracted from the two video soundtracks. A significant peak is found at the lag of around 24 seconds. The ENF signals from the two video recordings after alignment and the ENF measured from the power mains at the corresponding time are plotted in Fig. 2 (b). We observe the variation patterns of these signals match well with each other. Several video frame pairs after alignment are shown in Fig. 2 (c).



(a) (b)

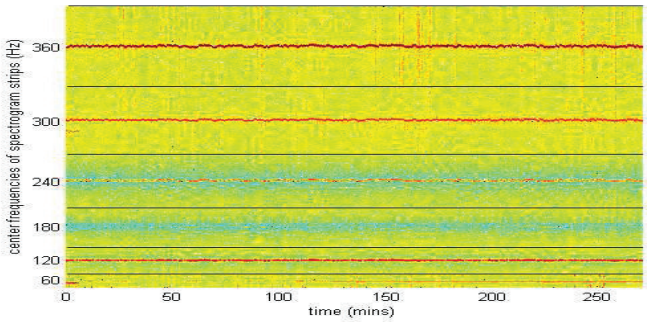
(c)

Fig. 2. Example of video synchronization by aligning the ENF signals.

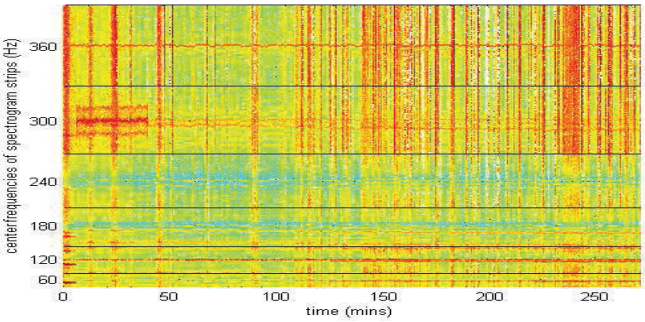
4. ENF FOR SYNCHRONIZING HISTORICAL RECORDINGS

Although most demonstrations of ENF being picked up by digital audio and video recordings in areas of electrical activities were reported in the recent decade, the presence of ENF can be found in analog recordings made throughout the second half of the 20^{th} century. For example, in our recent work, we demonstrated that ENF traces can be found in digitized versions of 1960s phone conversation recordings of President Kennedy in the White House [11]. Using ENF to analyze historical recordings can have many useful applications for forensics and archivists. For instance, many 20^{th} century recordings are important cultural heritage records, but some lack necessary metadata, such as the date and time of recording. Also, the need may arise to timestamp old recordings for investigative purposes, and ENF may provide a way to do that.

In this section, we explore aligning historical recordings temporally. We analyze two recordings from the 1970 NASA Apollo 13 mission [12] that we know were recorded at approximately the same time. The first recording is from the PAO (Public Affairs Officer) loop, which is the space-to-



(a)



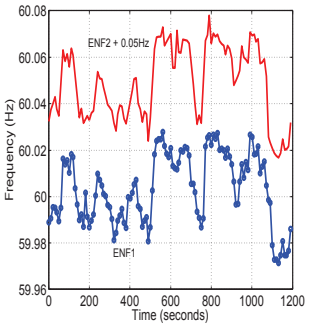
(b)

Fig. 3. Spectrogram strips around the ENF harmonics for the Apollo 13 recordings. (a): PAO recording; (b): GOSS recording.

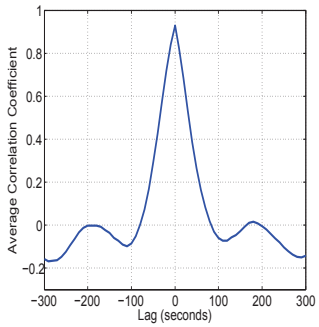
ground communications that was broadcast to the media. The second recording is of GOSS Net 1 (Ground Operational Support System), which is the recording of the space-to-ground audio as the people in mission control heard it. Both recordings are around four hours long. Figure 3 shows spectrogram strips for both recordings about the ENF harmonics. We can see that for the first recording, the ENF clearly appears around all the harmonics, and especially strongly around 360Hz. For the second recording, the ENF is noisier and it appears best around 120Hz and 360Hz.

We extract the ENF of the first recording from around 360Hz. For the second recording, we use the spectrum combining technique for ENF estimation [13], where we combine the ENF traces from around 120Hz and 360Hz to arrive at a more reliable ENF estimate. The resulting ENF signal is still rather noisy; we clean the signal by locating outliers and replacing them using linear interpolation from surrounding ENF values. Figure 4 (a) shows 20-minute simultaneous ENF segments from both recordings, with the second ENF signal displaced by 0.05Hz to be able to distinguish them and see them separately. Visually, the two signals look very similar.

In a synchronization scenario, we would need to match ENF segments from two or more signals with potentially different lags, and decide on the correct lag based on how similar the segments are, using the correlation coefficient as a



(a)



(b)

Fig. 4. Synchronize the Apollo 13 mission recordings with the ENF signals.

metric. As a proof-of-concept for the Apollo data described above, we divide the first Apollo ENF signal into overlapping 10-min ENF segments, and for each segment, we correlate it with equally-sized segments from the second Apollo ENF with varying lags. Since the two signals were recorded at the same time, this ground truth suggests that the highest correlation should be at zero lag. Figure 4 (b) shows the mean values of the correlations achieved for different lags, and we can clearly see that the highest correlation is achieved for zero lag which matches the ground truth.

We can see that the techniques discussed earlier for audio and video alignment can be extended to aligning two historical recordings of interest. This can potentially help timestamp old recordings of unknown date of capturing. With old recordings, we may not always have access to reference power ENF, as in the case considered here, yet we have the potential to utilize historical recordings of known date and time to create an ENF database to which we can compare recordings of interest that have uncertain information about capturing time.

5. CONCLUSION

In this work, we have explored the potential of the ENF signal for multimedia signal synchronization. The proposed approach works by extracting and aligning the ENF signals embedded in audio and video recordings. We have demonstrated our method with two applications: multi-view video synchronization and alignment of historical audio recordings. The ENF based synchronization approach has been shown to be effective, and has the potential to address challenging scenarios and complement other existing methods.

Acknowledgement This work is supported in part by NSF grants #1008117 (University of Maryland ADVANCE Seed Research Grant), #1309623 and #1218159.

6. REFERENCES

[1] C. Grigoras, "Applications of ENF criterion in forensics: Audio, video, computer and telecommunication analysis," *Forensic Science International*, vol. 167(2-3), pp. 136–145, April 2007.

[2] R. Garg, A. Varna, and M. Wu, "Seeing ENF: natural time stamp for digital video via optical sensing and signal processing," in *19th ACM International Conference on Multimedia*, Nov. 2011.

[3] M. Huijbregtse and Z. Geradts, "Using the ENF criterion for determining the time of recording of short digital audio recordings," in *International Workshop on Computational Forensics (IWCF)*, Aug. 2009.

[4] R. W. Sanders, "Digital authenticity using the electric network frequency," in *33rd AES International Conference on Audio Forensics, Theory and Practice*, June 2008.

[5] D. Rodriguez, J. Apolinario, and L. Biscainho, "Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," *IEEE Transactions on Information Forensics and Security*, vol. 5(3), pp. 534–543, Septemper 2010.

[6] A. Hajj-Ahmad, R. Garg, and M. Wu, "ENF based location classification of sensor recordings," in *IEEE Int. Workshop on Info. Forensics and Security (WIFS)*, Nov. 2013.

[7] R. Garg, A. Hajj-Ahmad, and M. Wu, "Geo-location estimation from electrical network frequency signals," in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013.

[8] A. Hajj-Ahmad, R. Garg, and M. Wu, "Instantaneous frequency estimation and localization for ENF signals," in *APSIPA Annual Summit and Conference*, Dec. 2012.

[9] "The first men on the moon: The apollo 11 lunar landing," http://www.firstmenonthemoon.com/.

[10] O. Ojowu, J. Karlsson, J. Li, and Y. Liu, "ENF extraction from digital recordings using adaptive techniques and frequency tracking," *IEEE Transactions on Information Forensics and Security*, vol. 7(4), pp. 1330–1338, August 2012.

[11] H. Su, R. Garg, A. Hajj-Ahmad, and M. Wu, "ENF analysis on recaptured audio recordings," in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013.

[12] "Apollo 13 audio recordings," https://archive.org/details/Apollo13Audio.

[13] A. Hajj-Ahmad, R. Garg, and M. Wu, "Spectrum combining for ENF signal estimation," *IEEE Signal Processing Letters*, vol. 20(9), pp. 885–888, September 2013.