Investigation of Electric Network Frequency for Synchronization of Low Cost and Wireless Sound Cards

Oleg Golokolenko
Technische Universität Ilmenau,
98693 Ilmenau, Germany
Email: Oleg.Golokolenko@tu-ilmenau.de

Gerald Schuller
Technische Universität Ilmenau,
98693 Ilmenau, Germany
Email: Gerald.Schuller@tu-ilmenau.de

***Abstract*—An electric network frequency (ENF) signal can be found in multimedia recordings due to propagation from the power grid. There a variety of applications based on ENF signal use, such as video and audio stream synchronization, and origin determination of multimedia recordings. In this paper, we propose the use of ENF to synchronize real-time audio streams from different, non-synchronized sound devices, for instance from wireless sound cards or low cost USB sound cards. Synchronization of separate audio streams can be achieved by aligning their embedded ENF signals. Our goal is to find out how accurate a synchronization using ENF at different SNR levels can be. We show simulation and real-time experimental results of audio stream synchronization. We found that with sufficient ENF level we can achieve an accuracy of the estimated delay difference of four samples at 44.1 kHz sampling rate.**

***Index Terms*—ENF, audio streams synchronization, low cost and wireless sound cards**

I. Introduction

Microphone arrays are widely used, for instance for source separation or for the attenuation of noise or unwanted sound sources. Usually they are part of a device with a single central clock, for instance a sound card with many microphone inputs. This might not be feasible or expensive when we have a distributed microphone sensor network, for instance an ad-hoc network consisting of several smartphones, or several sensor nodes with microphones, like microcontrollers with microphones and wireless connections which are distributed over some area. Our application example is an inexpensive and scalable microphone array which consists of several low cost USB-soundcards, where each has its own clock. Each specific USB-soundcard has its own associated sample buffer, to buffer slight sampling rate differences to the central processing computer. These buffers fill up with different speeds, and hence we obtain audio signal delay differences between the different sound cards. Our overall goal is to synchronize the real-time audio streams of several soundcards or several wireless devices by estimating these delay differences, to be able to apply array signal processing.

A possible solution to synchronize audio streams from several audiocards is to use a synchronization or time stamp signal, or a ”word clock”, at the level of the output of the A/D converter and before the buffer. But this would require specialized hardware. Instead, we try to achive the synchronisation by using a common signal which is available from the environment, the Electric Network Frequency (ENF). Since the instantaneous value of ENF fluctuates around its nominal value it is in princible even possible to use it for synchronization purposes even if the delay is longer than one period [1], [2] (a property which we don’t use here). A low level of the 50 or 60 Hz ENF is present in most audio streams, and it can be recovered using a suitable filter. This ENF signal results from power cables throughout our buildings, who generate 50/60 Hz magnetic fields, which then induce low level 50/60 Hz voltages in out microphone cables. Another path of ENF coupling is through the power supply of the sound cards, if their power supply comes from our electric network. In that case, residual levels of 50/60 Hz reaches the A/D converters in the sound cards.

A previous application of the ENF was to synchronize video and audio streams [1], but there the required accuracy is much less than for our audio application. We like to obtain a measurement of the delay difference with an accuracy on the order of a few samples of our audio stream. This is challenging because of the narrow band nature of the ENF.

Our goal for this paper is to find out how accurate we can measure or estimate delay differences between different sound cards using the ENF signal. This is done by using a delay difference measurement obtained from wide-band acoustical noise as ”ground-truth”.

The rest of our paper is organized as follows. Section II reviews existing approaches of synchronization in the case of wireless devices. It also mentions algorithms based on the application of ENF. In Section III measuring delay differences using wideband noise and ENF is described. An accuracy evaluation is given in Section IV. Finally, Section V concludes our paper.

II. Review of Current Approaches in A/D Converter Synchronization

Many different approaches exist for the clock synchronization. This aspect is not only important for the synchronization

ISBN 978-0-9928626-7-1 © EURASIP 2017

of A/D converters, but also in wireless acoustic sensor networks [3].

Wu et al. in [4] proposed an approach which uses time stamps for receiver synchronization. Even though it is quite simple, it does not suit our problem because we would need specialized hardware for it.

Markovich-Golan et al. in [5] proposed a blind sampling rate offset estimation. They assume sufficient background noise and a perfect speech activity detector, such that the noise can be used for synchronization in speech pauses. But we would like to have a system which works for general audio signals, for instance music, not just speech.

A similar approach to the mentioned above was proposed in [6], in this paper only the sampling rate difference is investigated, but not delay differences.

ENF is also used for forensic purposes, such as [7], [8], [9], [10], [11]. One work explores usage of ENF for synchronization of audio and video [1]. But there the precision of the delay differences can be much less precise than for our target application.

III. Measuring Delay Differences using Wideband Noise and ENF

The ENF is embedded in audio recordings at its nominal frequency value and its harmonics. Fig. 1 represents the spectrogram of an audio recording with ENF. The strong presence of the ENF can be observed.

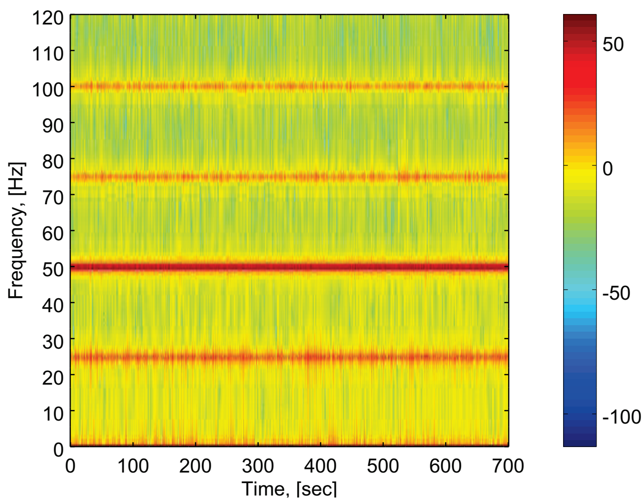

Fig. 1. Spectrogram of an audio signal with ENF.

Fig. 2 shows the 50 Hz ENF signal after narrowband 50 Hz filtering with the Laptop connected to network power. Observe that we clearly see a phase (delay) difference between the left and right microphone channel. The next Fig. 3 shows the ENF amplitude with the Laptop running on battery power. Here the amplitude is only about 1/20 of the case with network power. This shows that in this case most of the ENF signal is coming from the power supply.

Our test setup is 2 different USB sound cards, each with one microphone (a stereo setup). We used about 30 cm unshielded microphone cables (to pick up a sufficient amount of ENF) and electret microphones. The experiments were done in an office environment using three different USB sound cards with different pairings.

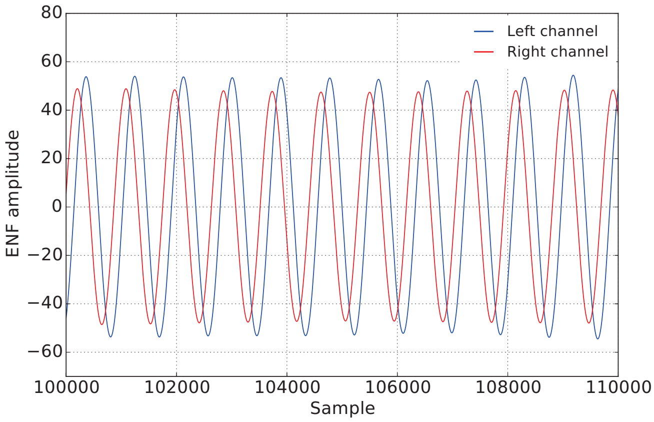

Fig. 2. Amplitude of the 50 Hz ENF signal (Laptop connected to network power), at 44.1 kHz sampling rate.

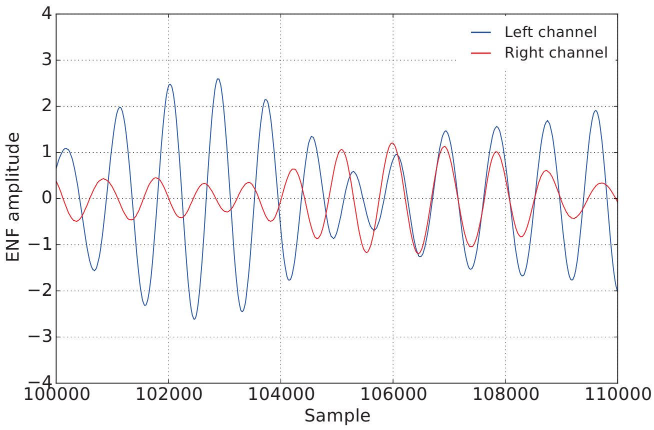

Fig. 3. Amplitude of the 50 Hz ENF signal (Laptop running on battery power), at 44.1 kHz sampling rate..

The positions of the microphones and speaker are shown in Fig. 4. The distance between the speaker and the microphones was 182 cm, because in that case the microphones will be in the far-field of the speaker. The speaker was used to play a uniformly distributed white noise, which was generated in software with Python. The diameter of the speaker was 33 cm. The speaker signal was used to obtain the "ground truth" for our delay differences. For real time processing and access to the sound cards we also used Python and its library "pyaudio".

To estimate the delay between two sound cards with white noise for our "ground truth", we found the position of the maximum of the cross correlation function between the 2 sound card signals. For a more efficient calculation we used the circular correlation, resulting from computation of the multiplication in the FFT domain.

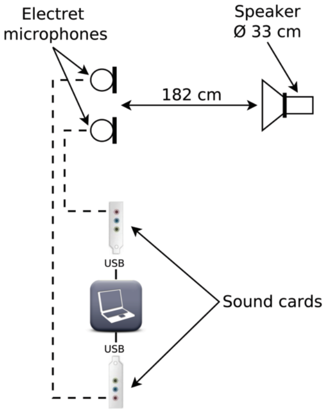

Fig. 4. Illustration of our experimental setup.

To obtain the ENF signal, we have used an IIR elliptic filter design in Python, using the function "scipy.signal.iirdesign", with the following parameters:

$$w_p = [\frac{49.0\ Hz}{f_s/2}, \frac{51.0\ Hz}{f_s/2}], w_s = [\frac{1.0\ Hz}{f_s/2}, 1]$$
$$g_{pass} = 5.0\ dB, g_{stop} = 110.0\ dB$$

where w_p and w_s are vectors of normalized passband and stopband edge frequencies (w_s looks somewhat artificial because it is basically the entire band without the passband, while avoiding a division by 0), f_s is sampling frequency and g_{pass} is the maximum loss in the passband, g_{stop} is minimum attenuation in the stopband. Frequency response of the filter is seen in Fig. 5.

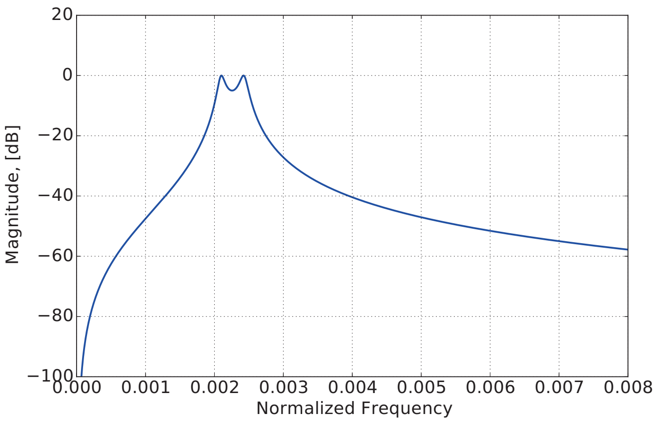

Fig. 5. Frequency response of the applied filter. The sampling frequency was $44.1\ kHz$.

We found that simply using the maximum of the (circular) correlation function gives us outliers for the wideband noise and noisy delay-difference estimates for the ENF case, as can be seen in Fig. 6. Our setup used a sampling frequency of 44.1 kHz and a block size (Chunk) of 1764 samples (chosen such that it contains 2 periods of a 50 Hz wave). We can see that clock drift follows a linear behavior, as expected.

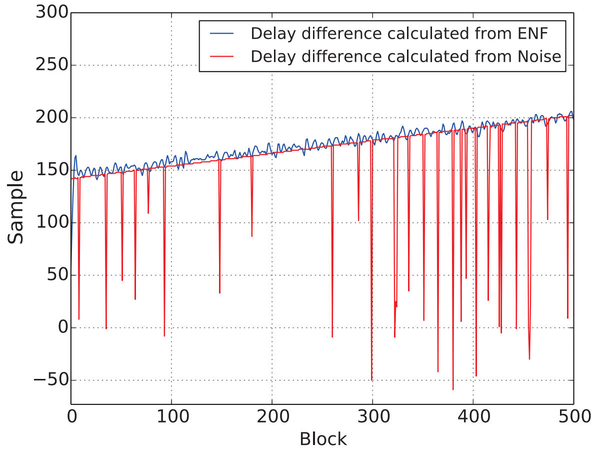

Fig. 6. Delay differences obtained from white noise and the ENF signal, from cross-correlation alone, for 500 blocks of 1764 samples each

The problem with the ENF signal is, that it has a much less pronounced peak in the cross correlation function. Hence we would like to find out how accurate our estimate for the delay difference is, and how we could make it more accurate. To improve the calculation of the delay difference, and make it more stable, we used additional techniques. First, we calculate delay difference through the FFT phase shift of our ENF signal. Then we combine the results from cross-correlation and the phase estimation, to obtain more stable results, and we fit a linear curve to our observations, to avoid outliers and reduce the influence of measurement noise. Such a combination is useful since none of approaches for delay difference calculation, neither cross correlation nor FFT phase shift, is perfect. The resulting delay difference tends to be more precise than all the components separately.

This results in Fig. 7 for the wideband noise case. Observe that we now indeed obtain a clean delay signal. For the different combinations of sound cards we see different drifts and offsets for the delay differences, as expected.

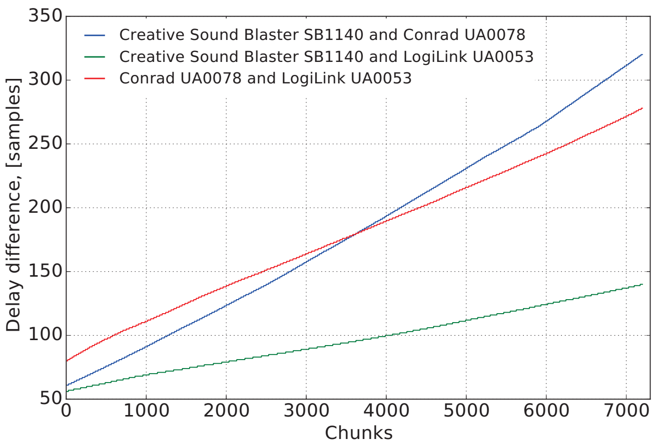

Fig. 7. Time plot of clock drift, observed for the different USB sound cards, using our improved estimation. The experiment duration was around 3 min.

ISBN 978-0-9928626-7-1

IV. Accuracy Estimation for the ENF Case

We can now apply our improved estimation techniques also to the ENF case, and see how it performs with different ENF strengths or different signal-to-noise rations in our ENF band (after our bandpass filtering). For that we start with simulated data. Then we compared it with real-time and real-world experiments.

A. Simulation setup and results

Used simulation data contains pseudo random noise, representing interfering signal of our wideband acoustic noise, and a 50 Hz sinusoid. Two channels with the same simulated data are shifted in time and passed into our proposed synchronization algorithm. The simulation was performed with different signal to noise ratio (SNR) to investigate the accuracy of the proposed synchronization algorithm. The results of two simulations, with +5 dB and -6 dB SNR, are shown in the form of cumulative distribution function (CDF) in Fig. 8.

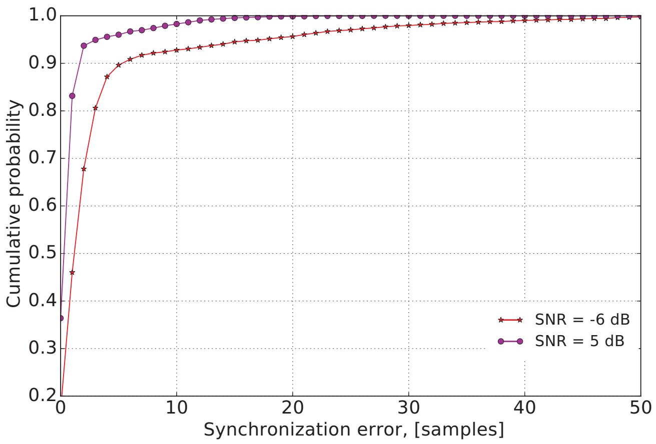

Fig. 8. Simulation results for ENF synchronization in the form of synchronization error CDF.

As can be seen in the simulation with -6 dB SNR synchronization, estimation errors of less than 5 samples appear in 90 % of the time. For the +5 dB SNR case we obtained about 2 sample accuracy 90 % of the time. During this experiment it was observed that SNR only slightly affects synchronization accuracy. Thus, one can conclude that a little presence of ENF can work properly with some degree of accuracy.

B. Experimental evaluation

To evaluate proposed synchronization technique, a real-time and real-world experiment was conducted. For this, the setup as shown in Fig. 4 was used. The Experiment was repeated 20 times with all possible combinations of USB sound cards. To compare results with simulation, real-time experiment was performed with -6 dB SNR. The results of real-time experiment can be seen in Fig. 9, again in the form of a CDF.

We can see that synchronization errors of less than 5 samples appear in 55 % of the time for 2 setups of sound

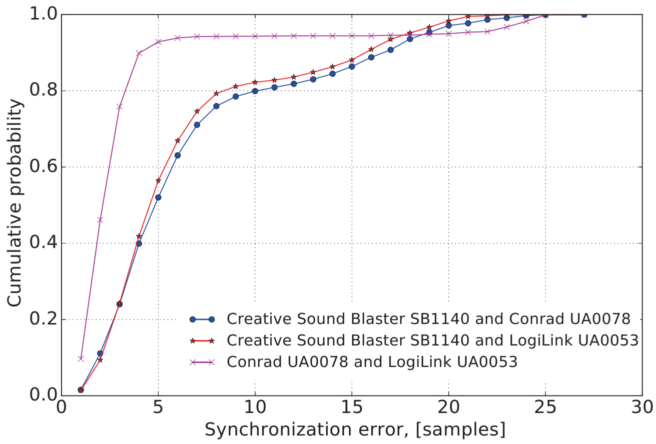

Fig. 9. Experimental results for ENF synchronization in the form of synchronization error CDF.

cards, and in 95 % of the time for the 3rd setup of sound cards. This strong dependence on the type of sound card was surprising, and might possibly be the result of internal noise or inaccuracies. The obtained overall average synchronization error was 4 samples. This results are worse than simulation possibly because the internal imperfections of sound cards, and the real ENF has a varying amplitude and frequency.

Moreover, the mean absolute timing error (MAE) was calculated to compare the achieved results in the real-time experiment against the state-of-the-art method [1]. In the case of current research, the MAE is about 0.14 ms, while in [1] it was 120 ms. Furthermore, audio block size used for synchronization was 0.04 seconds in our work and 16 seconds in [1]. In addition, proposed audio synchronization, as opposed to [1], was conducted in real time, which limits block size and demands synchronization accuracy of a few samples. This shows that we can improve to accuracies necessary for synchronizing audio streams.

Looking at Fig. 8 it might also be helpful to increase the level of the ENF signal to obtain better estimation accuracies for the delay differences. We can imagine little inductors at the microphones to boost the ENF level. The ENF signal could then later be removed using a notch filter. For beamforming applications the accuracy in samples reduces the maximum usable frequency accordingly.

V. Conclusions

We found that we can use the Electric Network Frequency (ENF) to estimate delay differences between audio stream with an accuracy of a few samples. Our improved estimation method helped to increase the estimation accuracy considerably. Artificially boosting of the ENF level helps to improve the estimation accuracy. This ENF signal can later be easily removed using a simple notch filter. Hence this represents a simple method to synchronize audio streams from non-synchronized audio sources, like different USB sound cards or wireless sound sources. Especially when it is not important to obtain, for instance, beamforming at full bandwidth.

REFERENCES

[1] Hui Su, A. Hajj-Ahmad, Min Wu, and D.W. Oard, "Exploring the use of ENF for multimedia synchronization," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 4613–4617.

[2] T. Baksteen, "The Electrical Network Frequency Criterion. Determining The Time And Location Of Digital Recordings," *Delft University of Technology*, November 11, 2015.

[3] N. D. Gaubitch Y. Zeng, R. C. Hendriks, "On clock synchronization for multi-microphone speech processing in wireless acoustic sensor networks," in *In Proc. IEEE Int. Conf. Acoustics, Speech, Signal Proc.(ICASSP)*, 2015.

[4] Yik-Chung Wu, Q. Chaudhari, and E. Serpedin, "Clock synchronization of wireless sensor networks," *Signal Processing Magazine, IEEE*, vol. 28, no. 1, pp. 124–138, Jan 2011.

[5] Shmulik Markovich-Golan, Sharon Gannot, and Israel Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*, Sept 2012, pp. 1–4.

[6] D. Cherkassky and S. Gannot, "Blind synchronization in wireless sensor networks with application to speech enhancement," in *Acoustic Signal Enhancement (IWAENC), 2014 14th International Workshop on*, Sept 2014, pp. 183–187.

[7] Sebastian Mann, Luca Cuccovillo, Patrick Aichroth, and Christian Dittmar, "Combining ENF phase discontinuity checking and temporal pattern matching for audio tampering detection.," in *GI-Jahrestagung*, Matthias Horbach, Ed. 2013, vol. 220 of *LNI*, pp. 2917–2927, GI.

[8] Ravi Garg, Avinash Varna, Adi Hajj-Ahmad, and Min Wu, "Seeing ENF: Power-signature-based timestamp for digital multimedia via optical sensing and signal processing," *IEEE Transactions On Information Forensics And Security*, vol.8, no. 9, pp. 1417–1432, September 2013.

[9] Robert C Maher, "Audio forensic examination," *Signal Processing Magazine, IEEE*, vol. 26, no. 2, pp. 84–94, 2009.

[10] Ling Fu, Penn N. Markham, Richard W. Conners, and Yilu Liu, "An Improved Discrete Fourier Transform-Based Algorithm for Electric Network Frequency Extraction," *Transactions on Information Forensics and Security, IEEE*, vol. 8, no. 7, JULY 2013.

[11] Luke Dosiek, "Extracting Electrical Network Frequency From Digital Recordings Using Frequency Demodulation," *Signal Processing Letters, IEEE*, vol. 22, no. 6, JUNE 2015.

ISBN 978-0-9928626-7-1 © EURASIP 2017