Finding Audio Geography using ENF

Indraprastha Institute of Information Technology, New Delhi, India

Puneet Jain, Prerna Singh, Abhinav Jadon, Ambuj Mehrish and A V Subramanyam

puneet13150@iiitd.ac.in, prerna13149@iiitd.ac.in, abhinav12122@iiitd.ac.in, ambujm@iiitd.ac.in, subramanyam@iiitd.ac.in

Finding Audio Geography using ENF

Puneet Jain, Prerna Singh, Abhinav Jadon, Ambuj Mehrish and A V Subramanyam
Department of Electronics and Communication Engineering
Indraprastha Institute of Information Technology, New Delhi, India

***Abstract*—Electric Network Frequency(ENF) is a recently developed technique for the authentication of audio signals. ENF gets embedded in audio signals due to electromagnetic interferences from power lines and hence can be used as a measure to find the geographical location and time of recording. Given an audio signal, this paper presents a technique of finding the power grid the audio belongs to. Towards this, we first extract the ENF sinusoid using a very narrow bandwidth filter centered around the nominal frequency. The filter is designed using a frequency response masking approach. The ENF is estimated from the ENF sinusoid using Short time Fourier Transform technique. In order to classify a given audio, we first estimate ENF from the ENF sinusoids obtained from audio and power signals. Further, we use a matched filter in order to decide the corresponding power grid of the audio.**

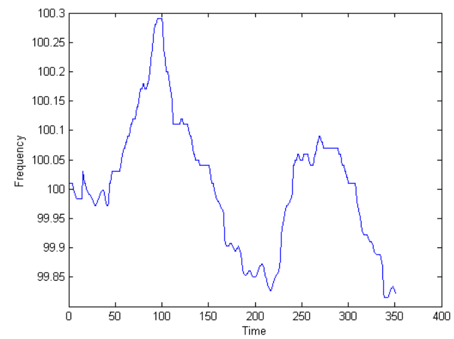

Fig. 1. ENF Values from audio

I. Introduction

Electric Network Frequency (ENF) is used for localization, authentication and forgery detection in digital audios. ENF gets embedded in the audio at the time of recording due to ambient electromagnetic activities. The comparison of ENF with the ground truth ENF of the power system can be used to predict the time and geographical location of the audio recording. In this paper,we describe the procedure of estimating the ENF from a given recorded signal. In addition, we uses a matched or correlation based detector to find the match of the audio signal to a power grid. The higher the correlation between the ENF of audio and power, it is more likely that the audio belongs to that audio grid.

II. Proposed Method

In this Section, we describe the various steps involved in the algorithm.

A. Demodulating and Filtering

In the proposed algorithm, we first design a bandpass filter to extract ENF sinusoid. Since the ENF sinusoid is a narrowband signal, a very narrow bandwidth, linear phase bandpass filter is desired. In addition, the desired filter should have sharp transition bandwidth and high stopband attenuation. Since the order of any linear phase filter is inversely proportional to its transition bandwidth, conventional filter design approach would lead to huge implementation complexity and may not be feasible for hardware implementation. In the proposed approach, filter with desired specifications is designed using a Frequency Response Masking (FRM) approach. The FRM approach is based on interpolation of prototype lowpass filter which results in multiband response. The resultant multiband response consist of multiple bandpass response with passband and transition bandwidth smaller by factor equal to the interpolation factor compared to the prototype lowpass filter. Hence, specifications of lowpass filter and interpolation factor need to be chosen carefully to obtain bandpass responses with desired specifications at required center frequency. Then, masking filter is used to mask undesired bandpass responses in multiband response. The use of FRM approach leads to at least 60% reduction in the number of multipliers making the proposed filter area and power efficient. Figure shows the filter response. From the Figure it is clear that the filter is very narrowband.

We use a filter at 120Hz for grids 1,3,9 and 100Hz for the others. We use different harmonics(original signal, 60Hz demodulated, 120Hz demodulated) of the signal to get a better estimation. We describe the estimation of ENF in the next section.

1) STFT: In the next step, we downsample the audio or power signal by a factor of two. As the highest ENF harmonic that we try to estimate has a nominal frequency of 240, the downsampling does not lead to an aliasing problem. Once we obtain the downsampled signal, we use FM demodulation to extract the ENF from different harmonics. We chose the following harmonics - 60 Hz, 120 Hz, 180 Hz, 240 Hz for 60 Hz power system and 50 Hz, 100 Hz, 150 Hz, 200 Hz for 50 Hz power system. We then use STFT technique proposed in [1] to estimate the ENF. The Figures 1 and 2 shows estimated ENF from one audio and one power signal.

Once the ENFs of all the audio and power were computed, we then investigated a matching algorithm. To determine if audio signal under investigation belongs to a specific grid, we

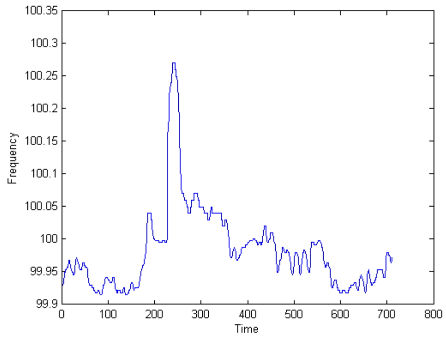

Fig. 2. ENF Values from power

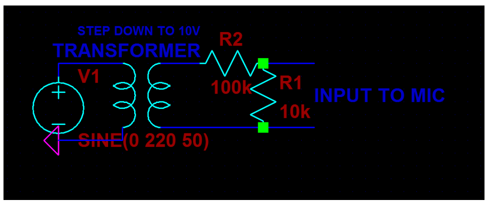

Fig. 3. Schematic of the Circuit

compute correlation between audio ENF and power ENF. The correlation will be high when the given audio is recorded in the corresponding grid from which reference power ENF is computed. In order to reliably decide whether the given audio belongs to a specific grid, we need to determine a threshold. Towards this, we tried to use a likelihood detector, however, we could not get a good threshold. And we resorted to analysing the maximum correlation values. Here, multiple harmonics of the ENF were extracted from both audio and power. These were correlated to give multiple values of correlation amongst an audio and power. In our analysis we find that the maximum 10 values of correlations corresponding to each audio signal, when correlated with all the power signals of the Training Dataset. We used them maximum values and the mode to analyse the signals. We set different thresholds for different Grids, according to their results from our estimation. These findings were also verified on the audio data collected in the presence of ENF interference in Indian power system.

B. Circuit

We created a simple circuit with a step-down transformer, from 220V to 10V, and a Voltage Divider to step it down further to 1V since the audio cards usually support 3V peak to peak. a schematic diagram and a snapshot of the circuit are shown in Figures 3 and 4. The power dissipated at the end of the mic (with the circuit) is 0.1mW(V = 1V, R = 10k).

III. Experiments and Analysis

We applied our algorithm on practice and testing dataset. We obtain the following results.

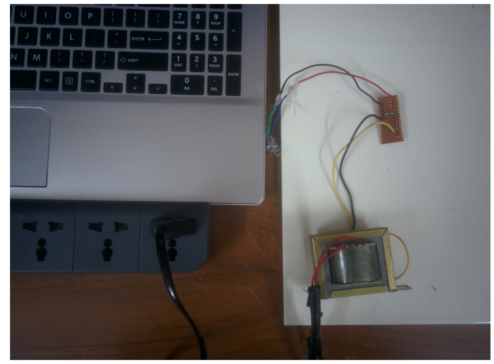

Fig. 4. Circuit deployed for recording

For Practice dataset -
AHCDN,NENNE,ADDDC,NCCDD,HDNCH,
CHDHE,DGNHG,DACGD,EHENG,DNGNC

For Testing dataset -
NGHND,DDEAE,DCNHC,HDDAH,HGEHH,
HHDFA,AHCEG,NCDND,HHCNN,HDHNA,
ENCNN,DHDHE,CACED,GDECE,CEDCN,
CNENE,AENHN,HADNC,DCCNN,EGNGD

The results are in the alphabetical order of the signals i.e. $Practice_1, Practice_2, Practice_3, ...Practice_{50}$.

For testing dataset - $Test_1, Test_2, Test_3, ..., Test_{100}$.

References

[1] A. J. Cooper, *The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordingsan automated approach*, Audio Engineering Society Conference: 33rd International Conference: Audio Forensics-Theory and Practice. Audio Engineering Society, 2008.