Environmental Noise Contamination Detector – Data Pipeline

Todd Schultz, Sean Miller, Rahul Birmiwal

Due: 21 Nov 2018

Data sources

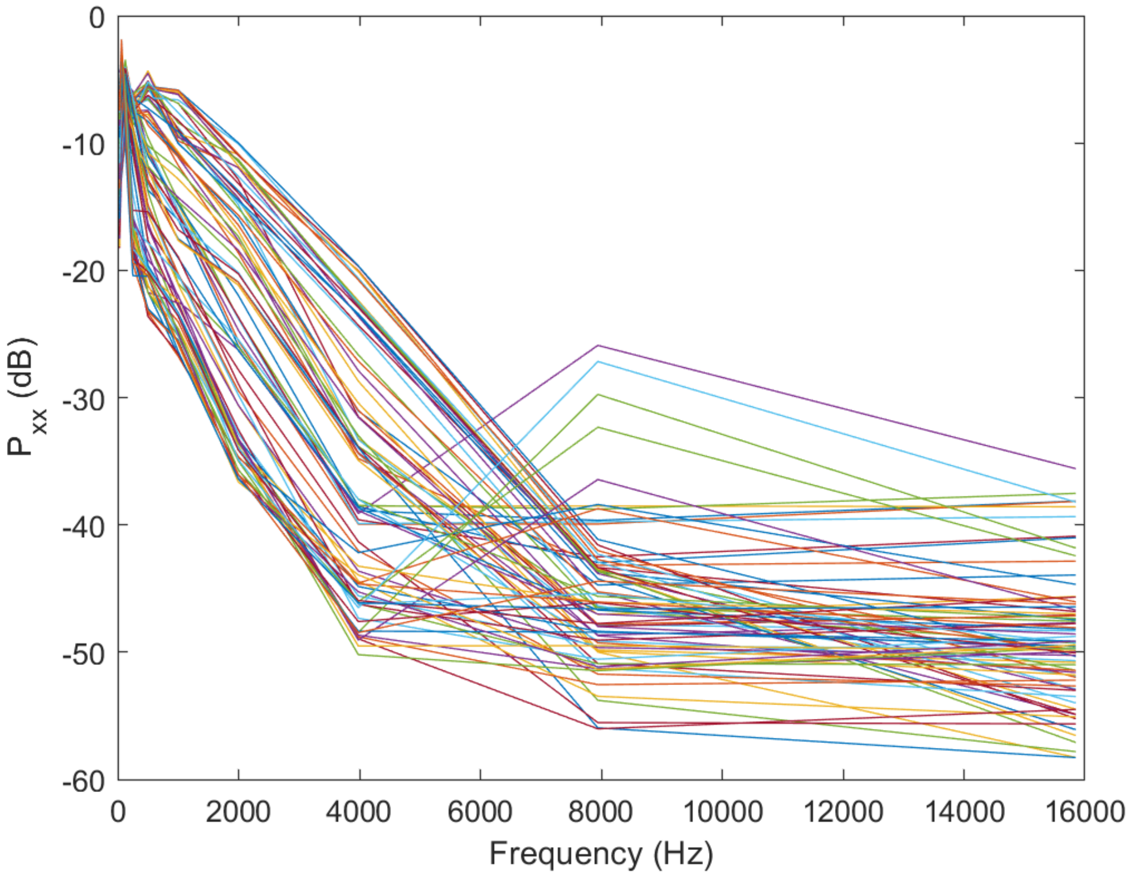
The data used for this project will originate from files. Typically, wave binary files, but the input data sources are not restricted to only this type of file. Since this is a research and development project, all data is expected to be in file format instead of streams or API calls. Extra data may be needed to expand the examples of wildlife noises.

We have 67 professionally recorded wave files. These vary in length from 10 seconds to a couple of minutes and feature aircraft noise signature, wildlife noises, and ambient recordings. Each recording was captured at a sampling frequency of 51.2 kHz. The signals will be broken down into smaller increments typically on the order of one to two seconds, thus each file will generate numerous feature sets. The files have been provided by the project contact to each team member to store locally as total size is over 500 MB and not suitable for storage in a GitHub repository.

All the data files are in the wave file format, which is easily processed by numerous applications. Any data that may be added later is expected to be in a file format which would be easily accessible in data science tools. We are planning on using MATLAB for all our processing.

Data example

An example of our data (the plane08.wav file of a turbofan jet aircraft) that has been imported and processed to normalized octave levels is presented below. This file contained 3,072,000 samples. The time series was split into 1 second records that were overlapped by 25% generating 79 records. The octave band power spectrum was computed for each record for the bands centered from 30 Hz to 16 kHz. The power levels of the octave spectrum were normalized such that the total power present in the spectrum is unity. The normalized octave spectrum will be one feature set used to test the identification of environmental noise.


0
-10
-20
-30
-40
-50
-60
P_{xx} (dB)
0
2000
4000
6000
8000
10000
12000
14000
16000
Frequency (Hz)