A note on the HUMS2023 Dataset

7-Mar-2024

From the results of the HUMS2023 Data Challenge, we observed that some participants have identified notable anomalies in a handful of data files associated with early-stage fault development, e.g. #15, #50, #125, #143 and #150 of the original dataset.

While these anomalies were not found to be related to the planet gear rim fault, we believe they were likely caused by some occasional measurement-related issues during data acquisition.

Upon inspecting the raw vibration and tachometer signals from these files, we located some unexpected prominent signal spikes likely related to extraneous electrical noise. While the presence of these spikes in the raw data is not favourable to analysts, spikes of this nature can be commonly encountered in many industrial field measurements due to issues in cables, sensors, amplifiers and various interferences to the data acquisition system.

It is worth noting that while the data (containing specific files with occasional measurement-related issues) can well represent real industrial vibration data, it is important to raise attention to these specific files.

This is particularly important for researchers using pure data science based techniques. Consequently, we have attempted to manually rectify these spikes and have regenerated the hunting tooth synchronous signal averages (H-SSA) in the following data files:

- #15, #50, #125, #143 and #150 in the original dataset for HUMS2023 Data Challenge.
- #15, #16, #71, #86, #90 and #262 in the additional dataset.

Note that this data rectification was conducted through local linear interpolations around the erroneous spikes to mitigate their effect in the raw data before computing the H-SSA's.

The only exception was file #86 in the additional dataset, where the raw tacho signal was observed to be heavily corrupted in two sections. For this specific file, the H-SSA was regenerated through a reference tacho derived from the ring front (RF) vibration channel (i.e. channel 2) using the phase demodulation method.

The fixed H-SSA's are appended to the two datasets for those researchers who intend to use the datasets. It is important to note that this rectification process may not have completely removed the effects of these erroneous spikes on the H-SSA's and that these effects could still surface during advanced stages of analysis.

The names of the two *mat* files and their contents are explained below:

- **For the original dataset:** *"xah_data_fixed_for_**DC**_files#15-50-125-143-150.mat*". This file contains five matrices, each containing 4 vectors corresponding to the fixed H-SSA's (1: Input Pinion, 2: Ring Front, 3: Ring Left, 4: Ring Rear) with raw data timestamps in each variable name in addition to the file sequence number of the dataset. For example, variable name of "xah_015_20211208_113917" means the fixed H-SSA for the 15th file in the original dataset with raw data recording date of 20211208 and time of 113917.

- **For the additional dataset:** "*xah_data_fixed_for_**added**_files#15-16-70-86-90-262.mat*". This file contains six matrices, each containing 4 vectors corresponding to the fixed H-SSA's (1: Input pinion, 2: Ring Front, 3: Ring Left, 4: Ring Rear) with raw data timestamps in each variable name of the fixed H-SSA's.