Exploratory Data Analysis – Fish Sonar Project

The dataset for this project has 14575 rows, and each row represents a singular fish from a particular species. In contrast, each column is labelled with an emitted sound frequency (ranging from 45 to 247) and consequently, each cell corresponds to the reflected signal strength (in decibels) from each fish for a particular sound pulse.

The overarching goal is thus to develop a predictive model that can identify a particular fish species given its reflected signals. Also, the models needs to be able identify significant associations between frequencies and specific species. For example, do trout have a unique signal at a frequency of 45 Hz? Do whitefish have a unique signal at 247 Hz? These are only examples of questions that we hope the predictive models can answer. However, for the preliminary data exploration aspect, the goal was to discover any trends between the frequencies and the reflected signals, hoping to identify differences between the three fish species. Note that each individual fish has repeated measurements in this dataset. This means that a single fish was 'blasted' with sonar from 45Hz to 247Hz a handful of times, and the reflected signal strength was recorded for each event. In order to visualize any potential trends, we manipulated the data accordingly:

- Filtered the original set by species and created 3 new datasets (one for each species)
- For each new data set we used the 'group_by' function to group the data set by its individual fish ID number, we then computed the average reflected signal strength for these 'groups'

- We then took the transpose of the aforementioned dataset and computed the average reflected signal strength: this computed the mean signal strength across a single frequency. For example, there were 21 individual trout's that were measured and each trout was 'blasted' with a sound pulse of 45Hz. The reflected signal from these blasts were recorded and we computed the average signal strength across the 21 individuals. Note that missing entires were removed from out calculations
- We now had three new data sets: one for each species. The rows represent the frequencies ranging from 45Hz to 247Hz and the variable of importance is 'Average_Frequency' (i.e the average signal strength across all the individual fish belonging to a particular species).

The results from these datasets are shown in Figures 1-9. In these figures, the average reflected signal strength for all three species are compared against one another for a range of frequencies. To begin with, the regions where the plots are identical to each other are trivial at this point considering we cannot differentiate between different species. Yet, in figure 2, there is a stark contrast between the three species along the 90-95 Hz range. The trout clearly have the lowest sound strength here with the bass having the second lowest and the whitefish the highest. This trend seemingly continues in figure 3 within the 95Hz to 119.5 Hz range but the discrepancies seem to dwindle ass the Hz increase. Another notable area of interest is shown in figure 6 within the 170 Hz to 197Hz range. In this plot, the trout now have the highest average sound strength while the whitefish have the second highest and the bass the lowest. The trout continue to have the highest sound strength in the rest of the figures (i.e. 197.5 Hz - 260 Hz), however, the averages between the bass and whitefish are almost identical.

Overall, we have identified two frequency ranges that could be critical in our future prediction model. The 90-95 Hz range and the 170-197Hz are seemingly the only two regions where all three species are clearly differentiable among each other with respect to their average reflected signal strength. Although these findings are preliminary in nature and lacking statistical theory, they do give us a visual representation of our data, compare the three species against one another, and finally show potential regions of interest that may influence our prediction model.

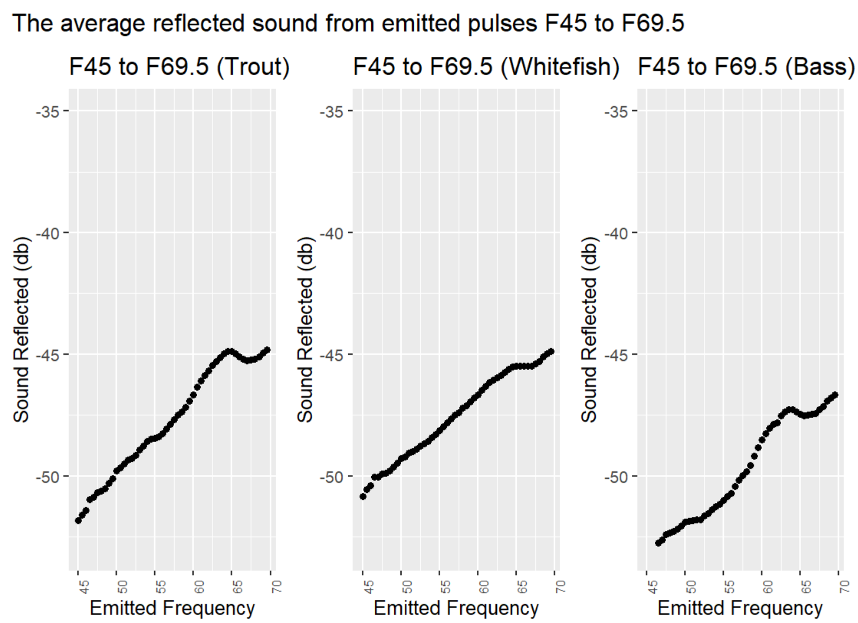

Figure 1: Comparing the reflected sound pulse from each fish between F45 - F69.5

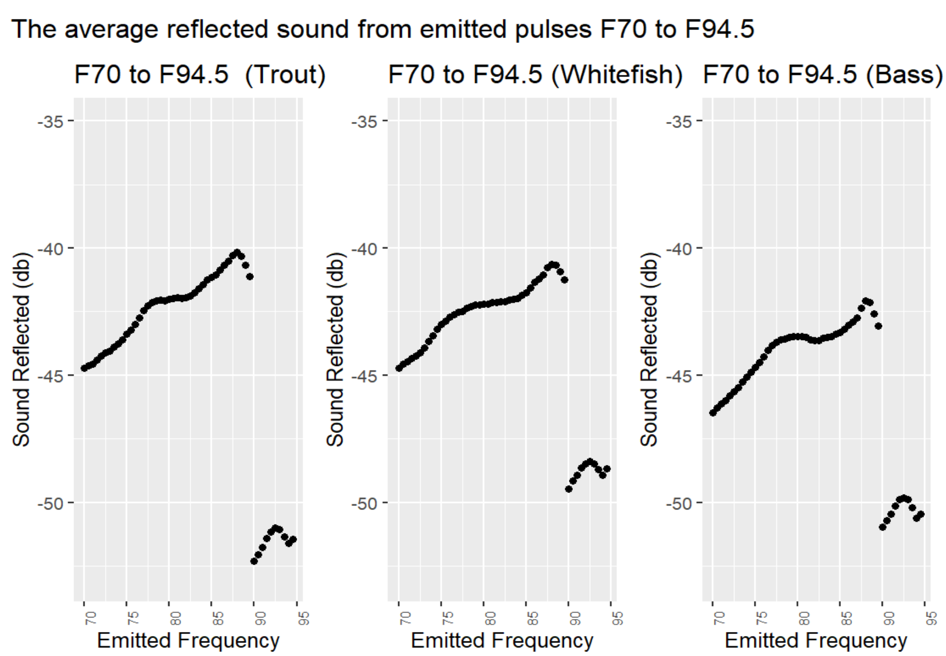

Figure 2: Comparing the reflected sound pulse from each fish between F70 - F94.5

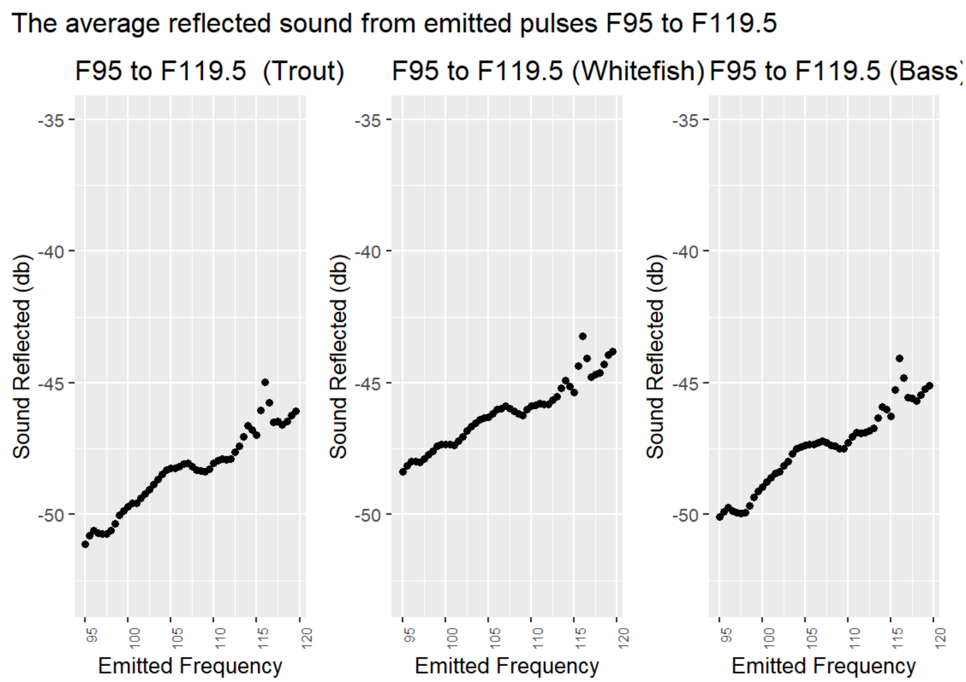

Figure 3: Comparing the reflected sound pulse from each fish between F95 - F119.5

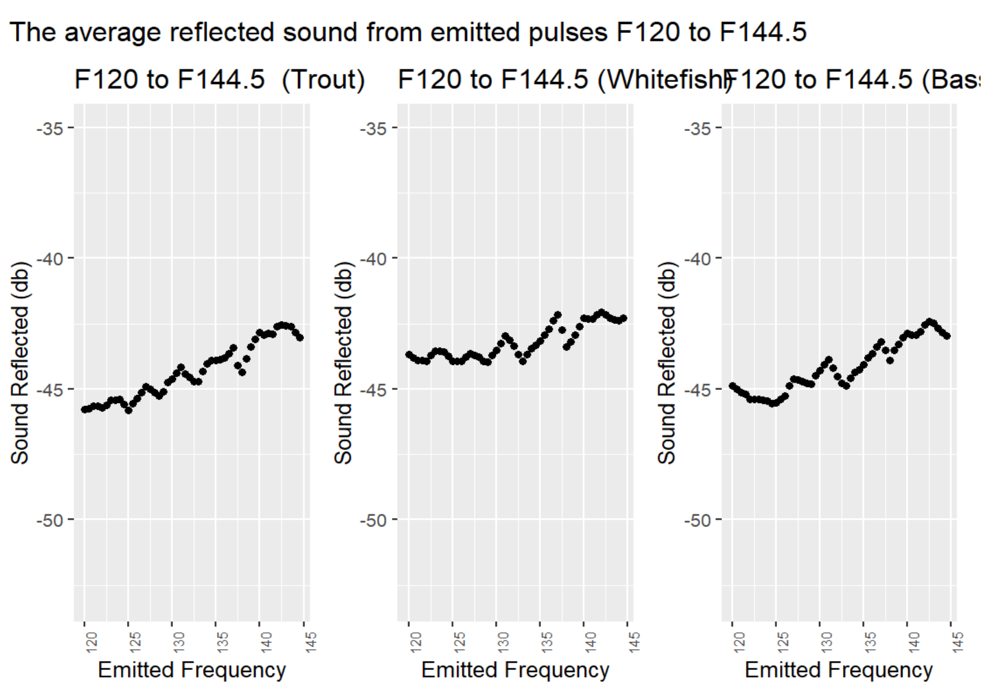

Figure 4: Comparing the reflected sound pulse from each fish between F120 - F144.5

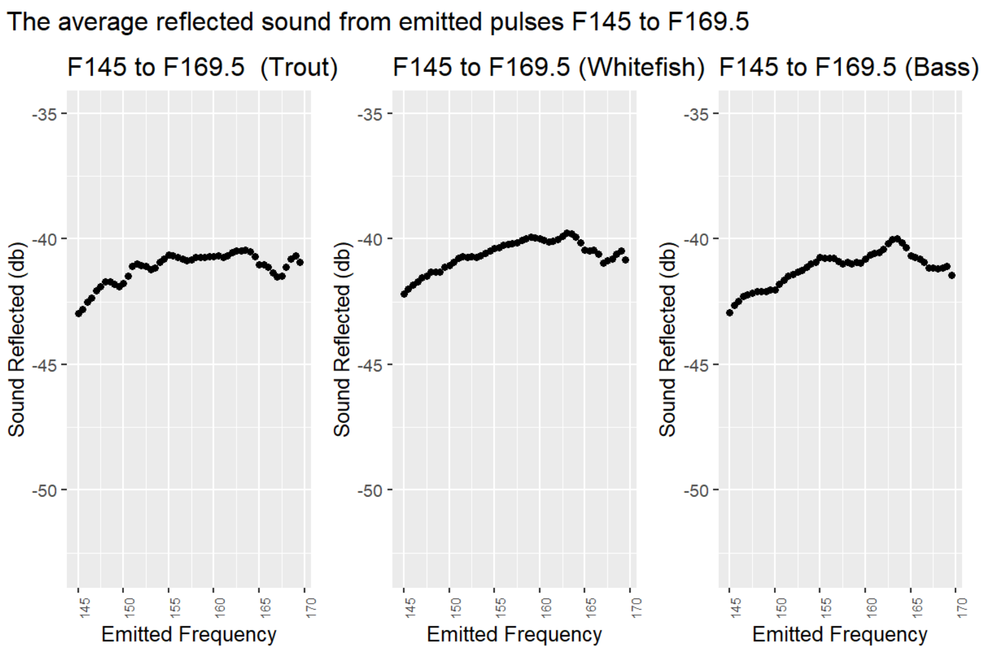

Figure 5: Comparing the reflected sound pulse from each fish between F145 - F169.5

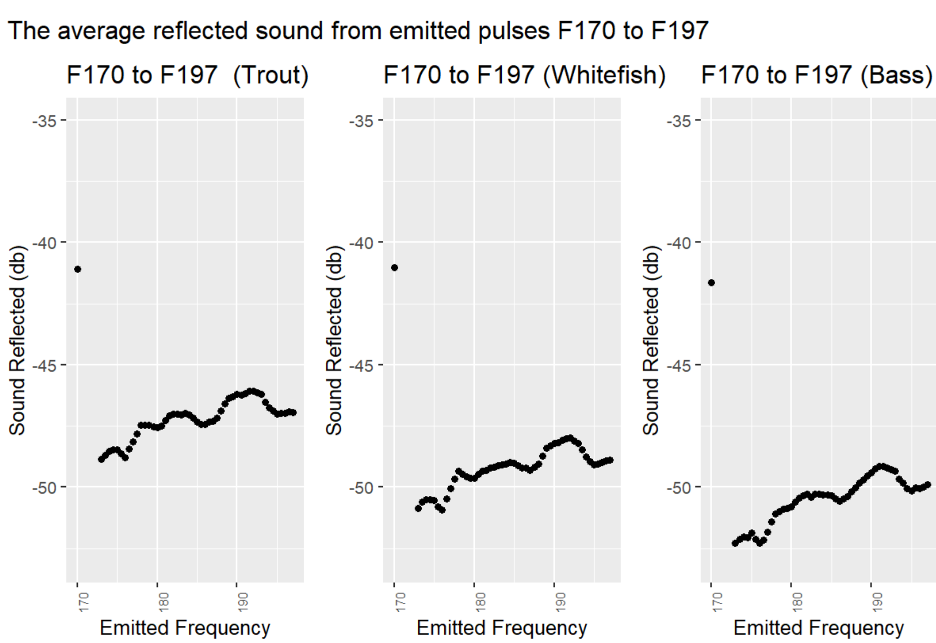

Figure 6: Comparing the reflected sound pulse from each fish between F170 - F197

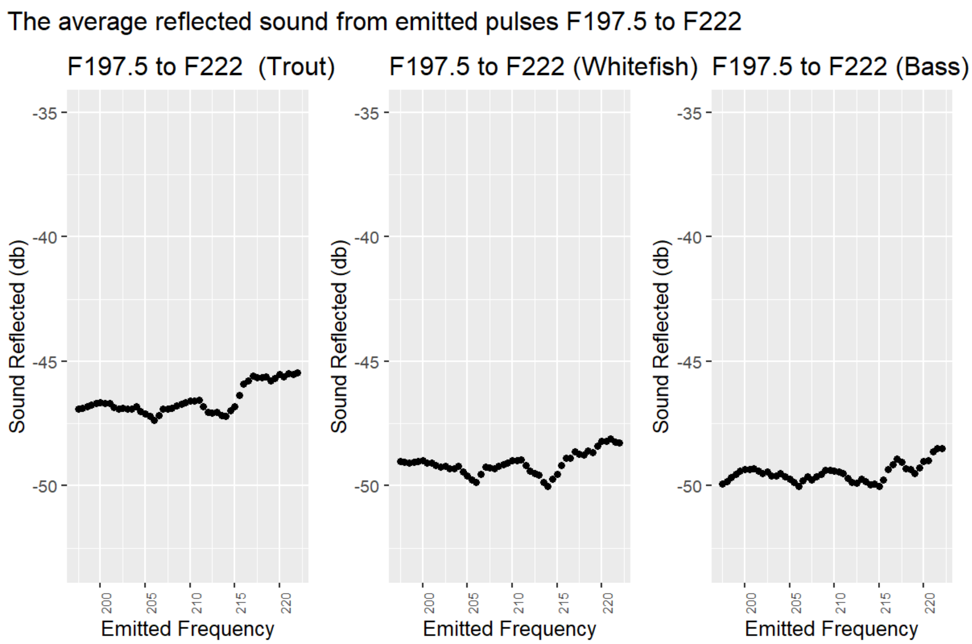

Figure 7: Comparing the reflected sound pulse from each fish between F197.5 - F222

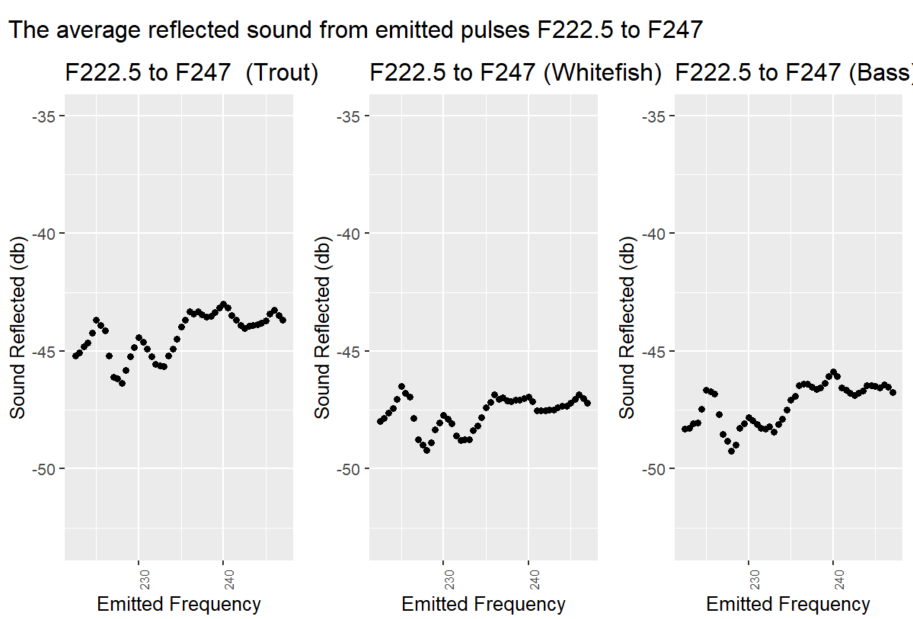

Figure 8: Comparing the reflected sound pulse from each fish between F22.5 - F247

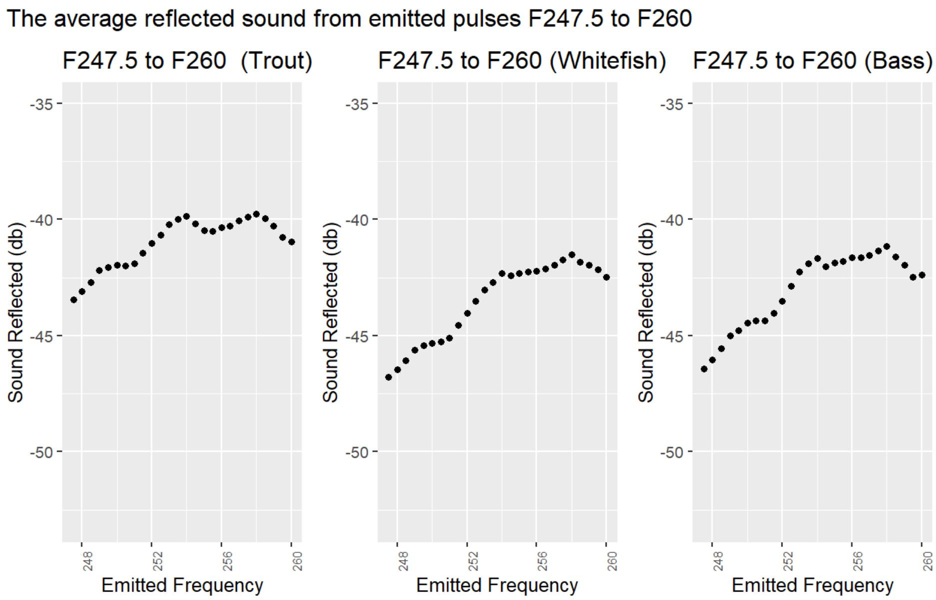

Figure 9: Comparing the reflected sound pulse from each fish between F247.5 - F260