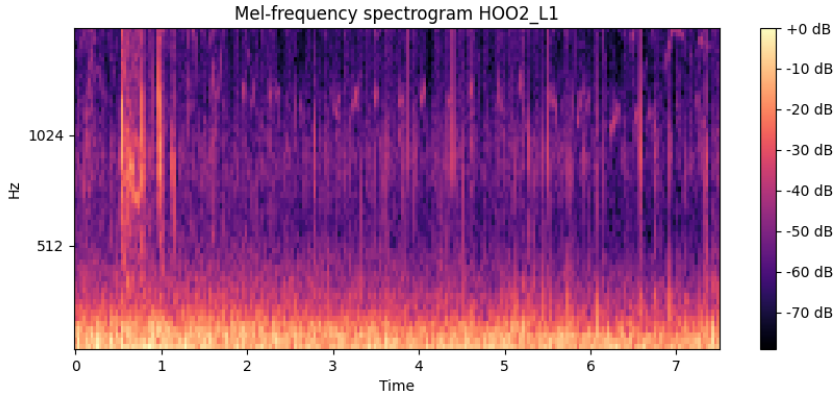
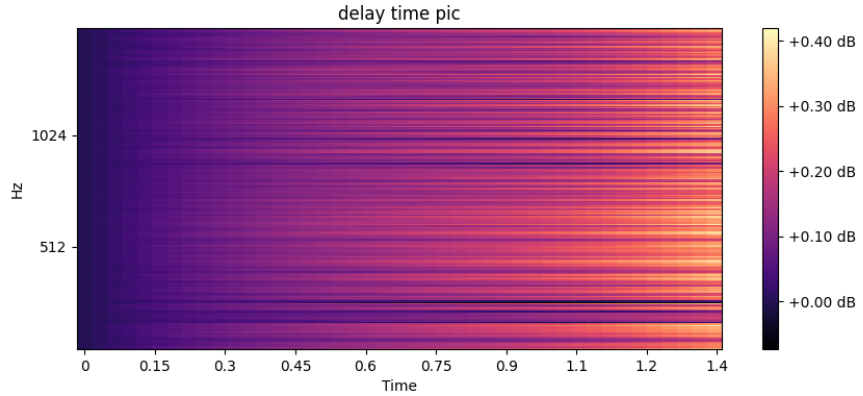
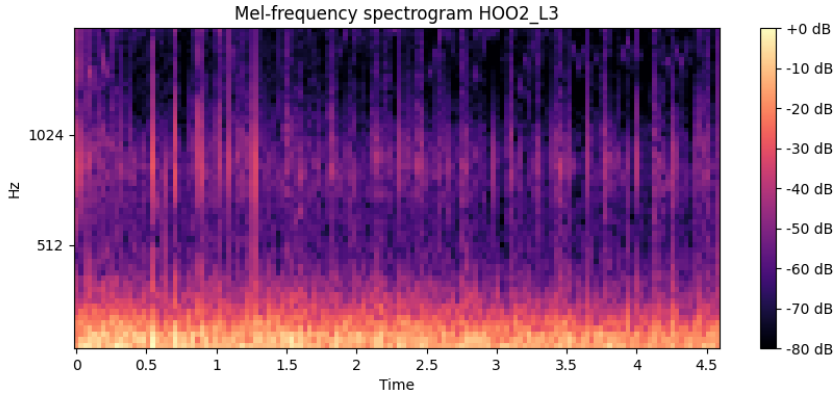
Delay Time Images

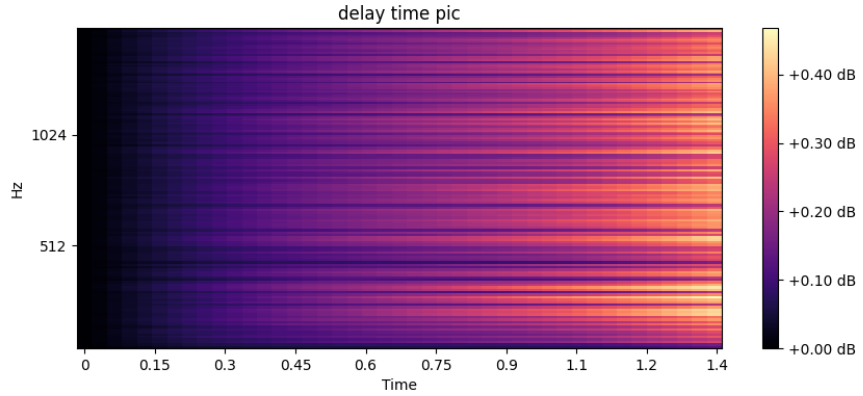
Owen Godsall Part III Project

06/02/2025

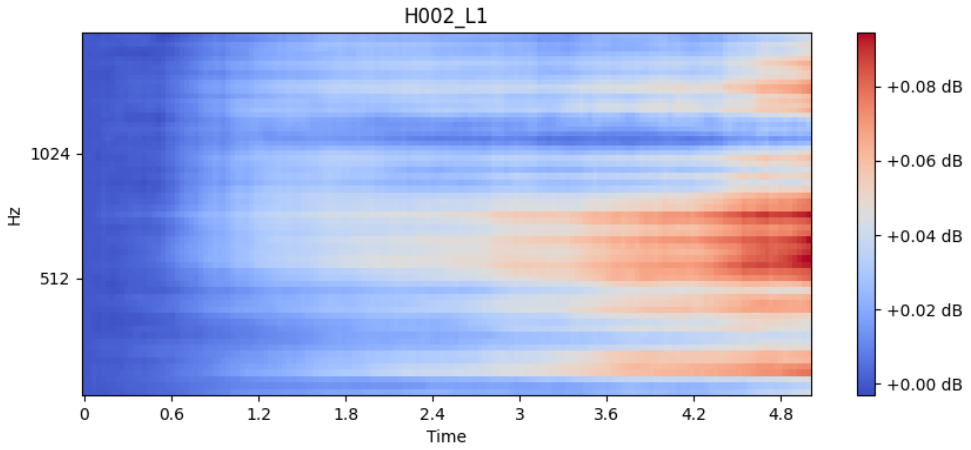


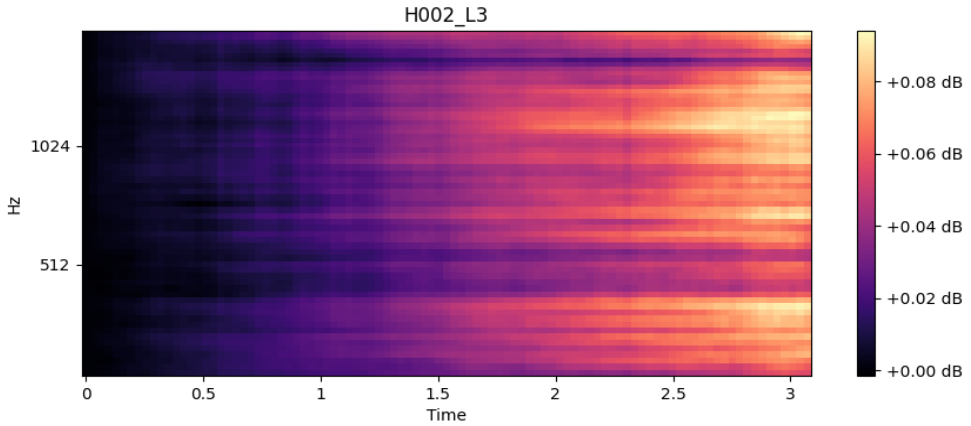






These delay plots to the right are wrong, they are of fixed time length which doesn't make sense considering the variation in length of audio clip. The 'pre_process_mel_spectrogram_t' function returns the transpose of the spectrogram, in these plots on the right, this isn't accounted for in analysis.



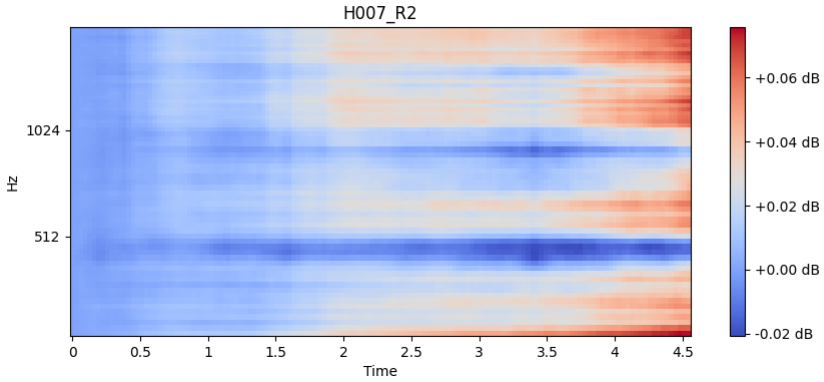


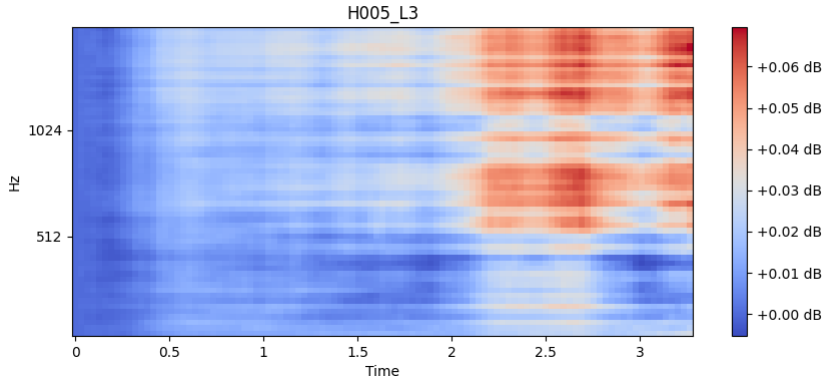
These are the plots produced (same audio clips as previous slide) when the transpose is corrected for.

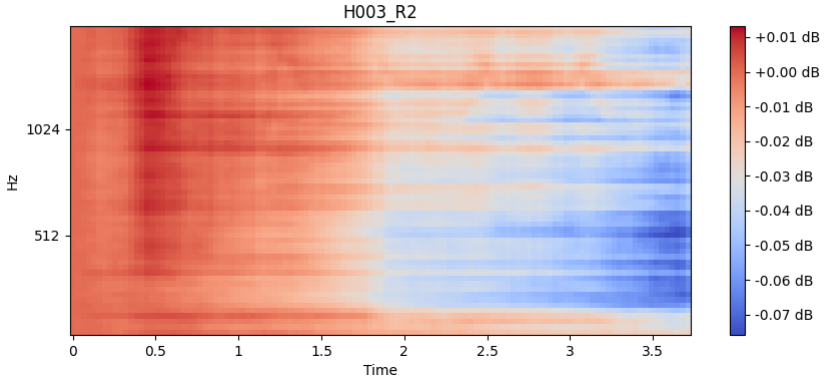
The length of the delay plots now matches up with the length of the initial audio clips. Can now also see more unique variation between audio clips.

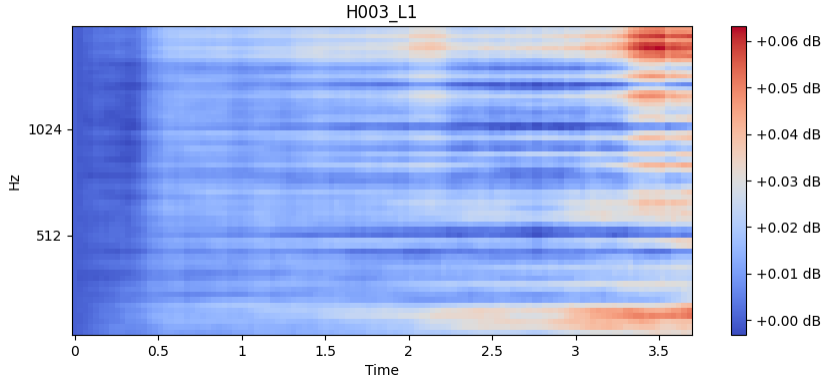
Still much shorter length of spectrogram than the length of the raw audio clip – 30s goes down to roughly 4.9s and 18s to roughly 3 (approximately reduced by 6x).

Examples of notable variation

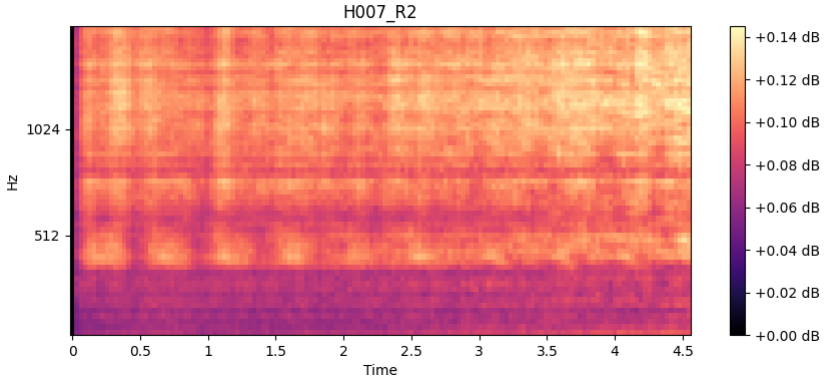


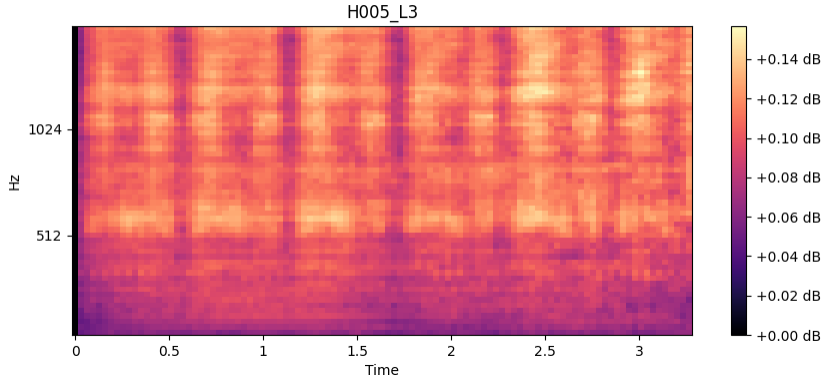


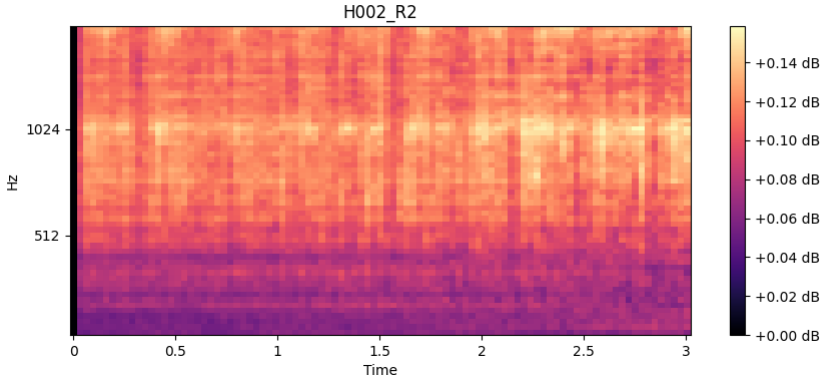


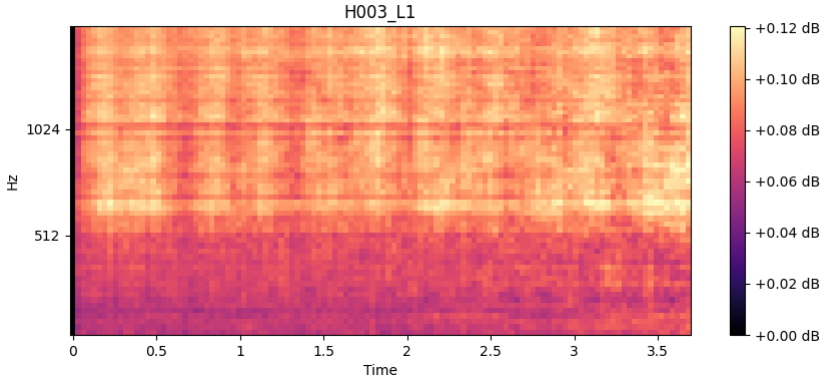


Altered images taking the modulus of the difference of snippets

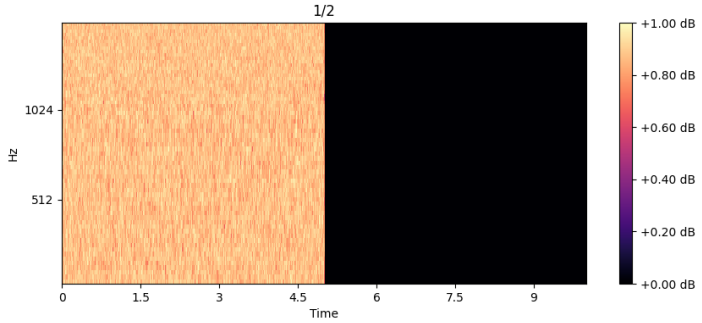


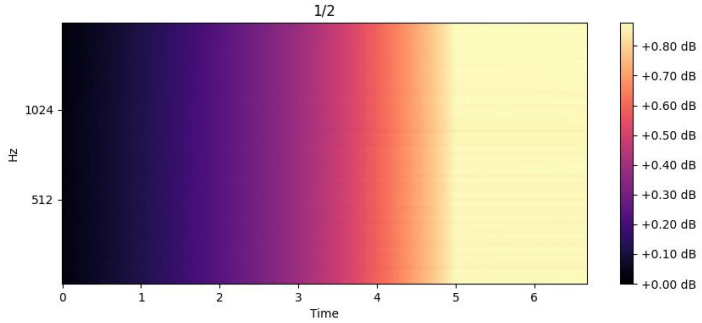


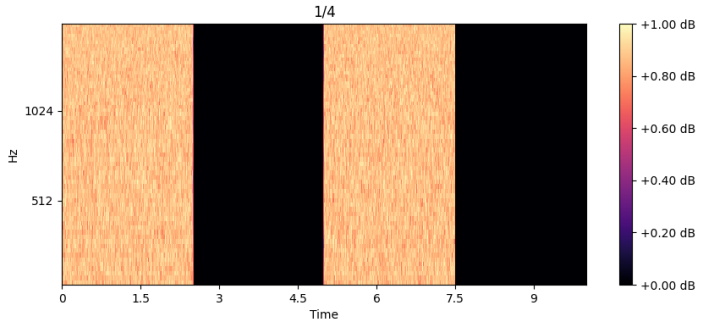


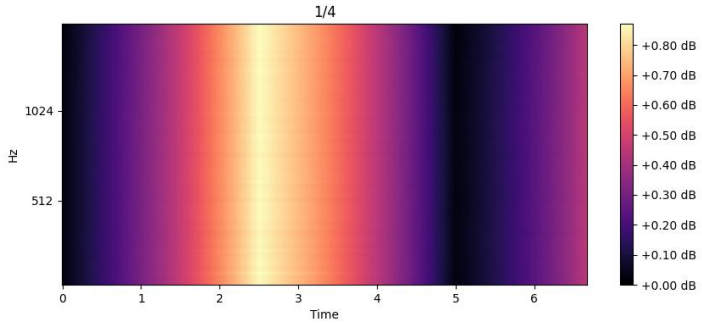


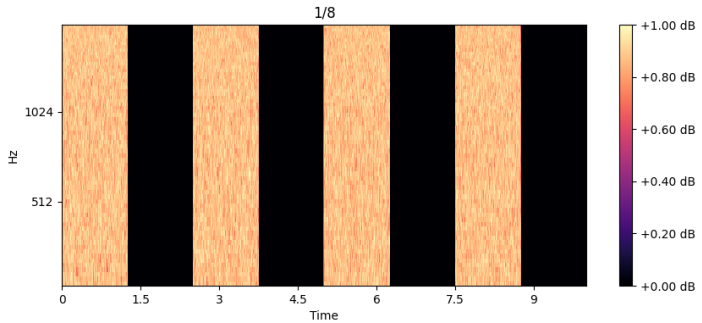
Example Pictures:

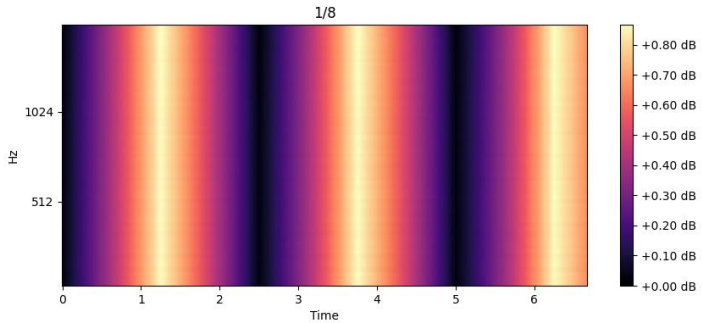








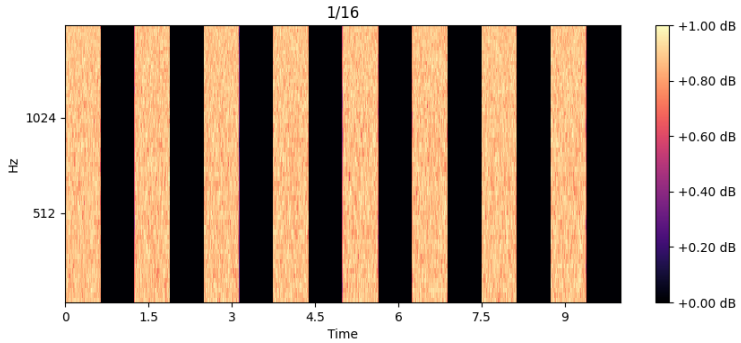


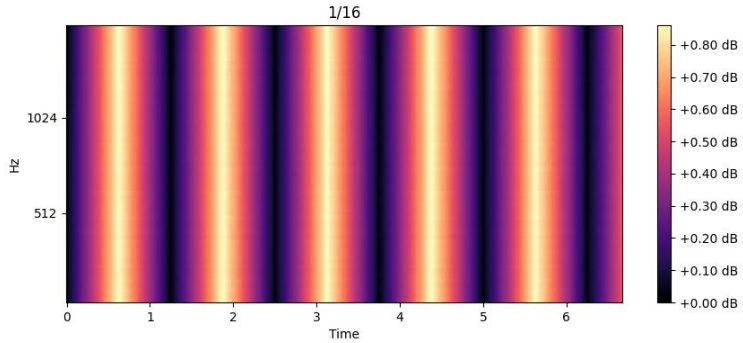


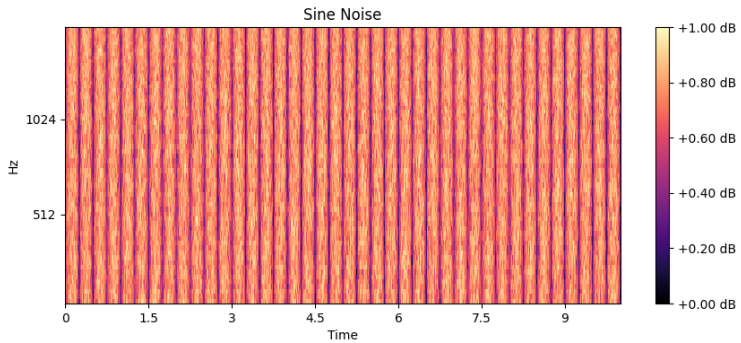
These examples demonstrate how the transformation to delay time in this way reduces the length of the clip.

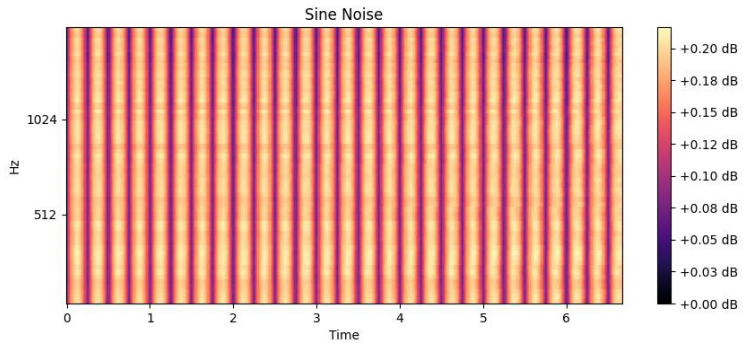
The length reduction is removing what is obvious in these cases would be continuation of the oscillating intensity trends.

Example Pictures:



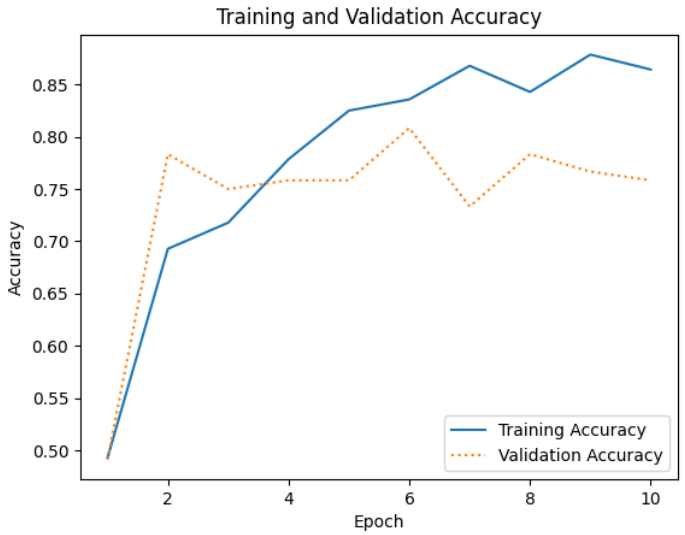


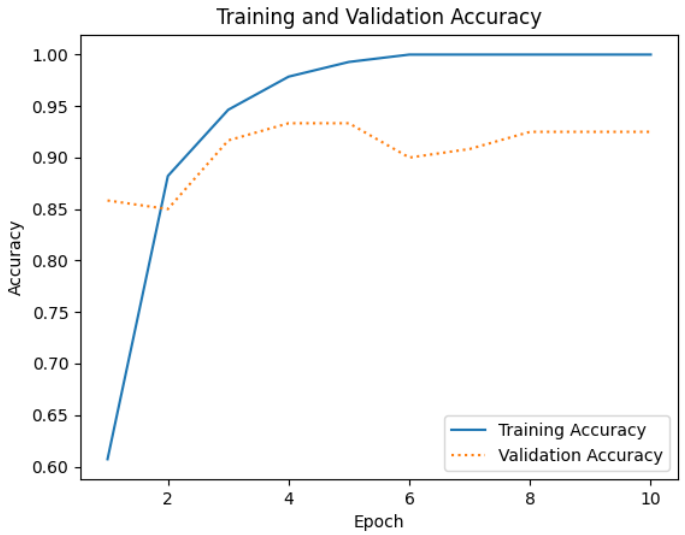


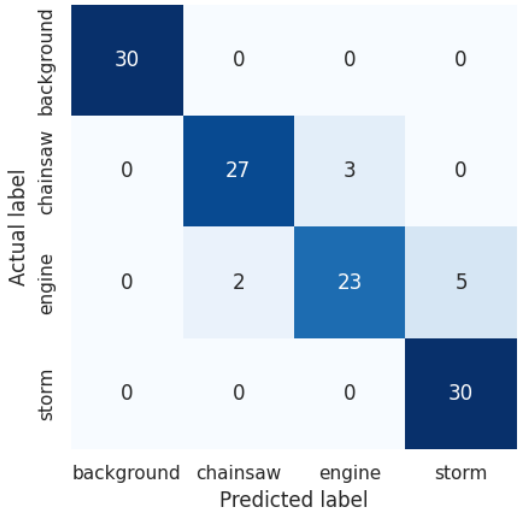


Sine goes to sine and frequency is preserved

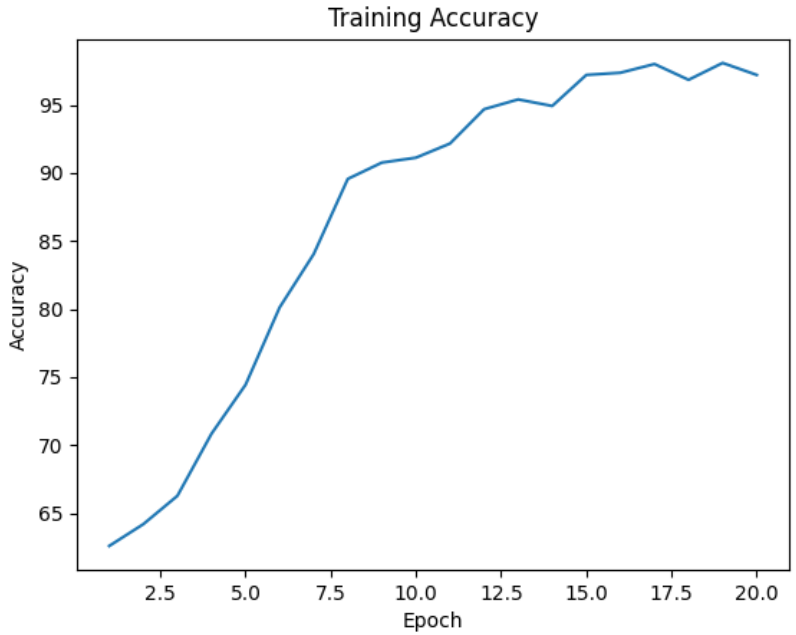
Logging Case Study:





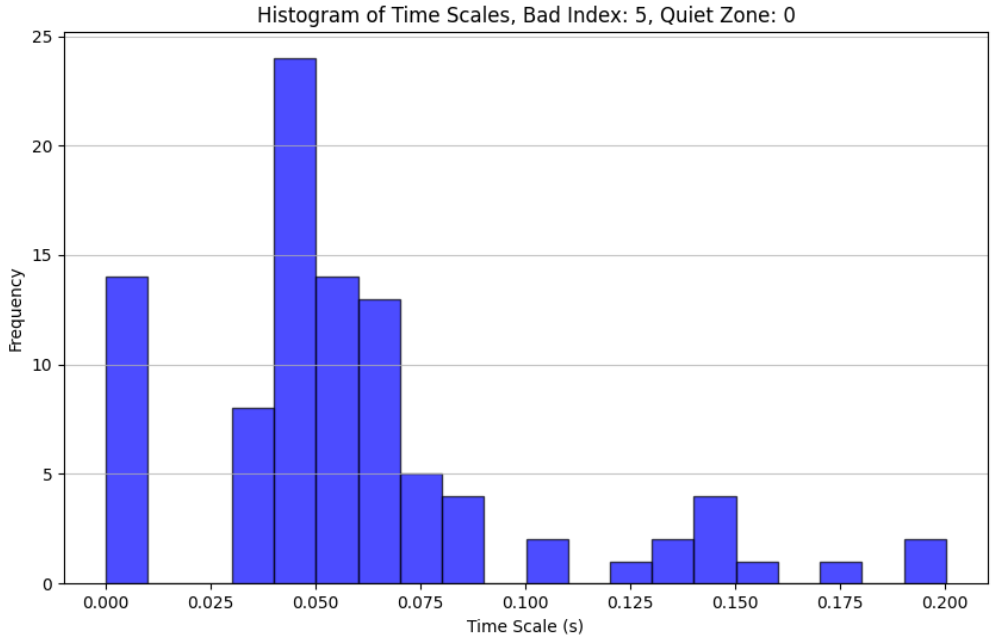


GoogLeNet Transfer Learning PyTorch:



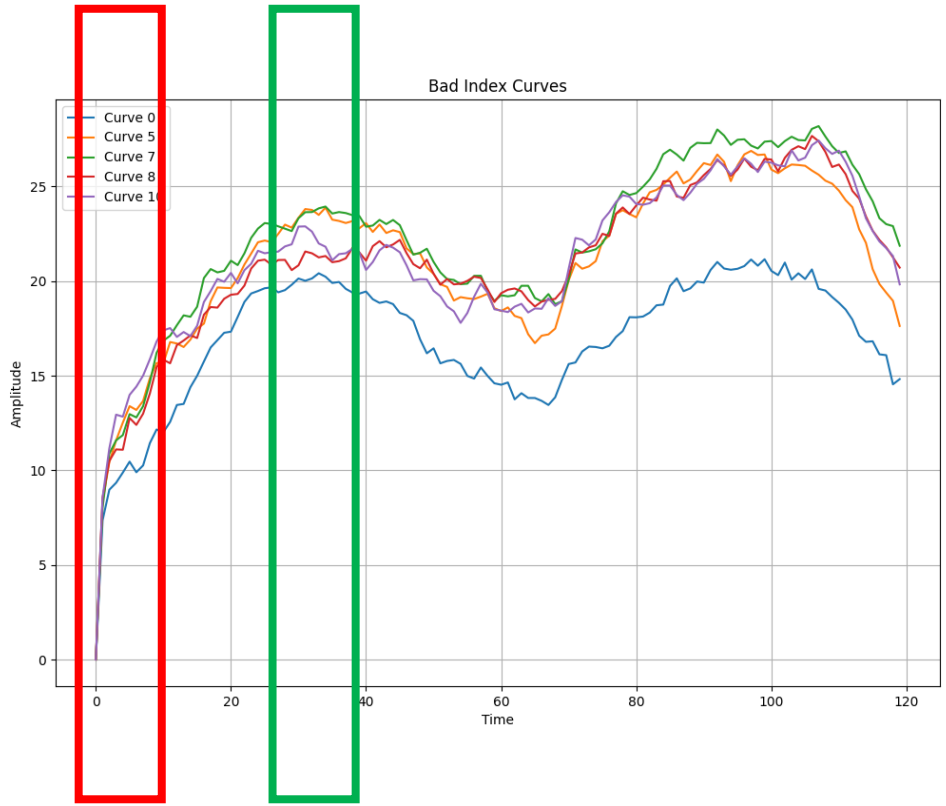
- Learning rate 1e-3
- Need to do a validation test
- Have been logging run times, not sure whether seeing as input dimensions are always the same I'll see appreciable speedup
- AUC of 0.98 on this is way too high

Exploration of Timescales of decorrelation onset:



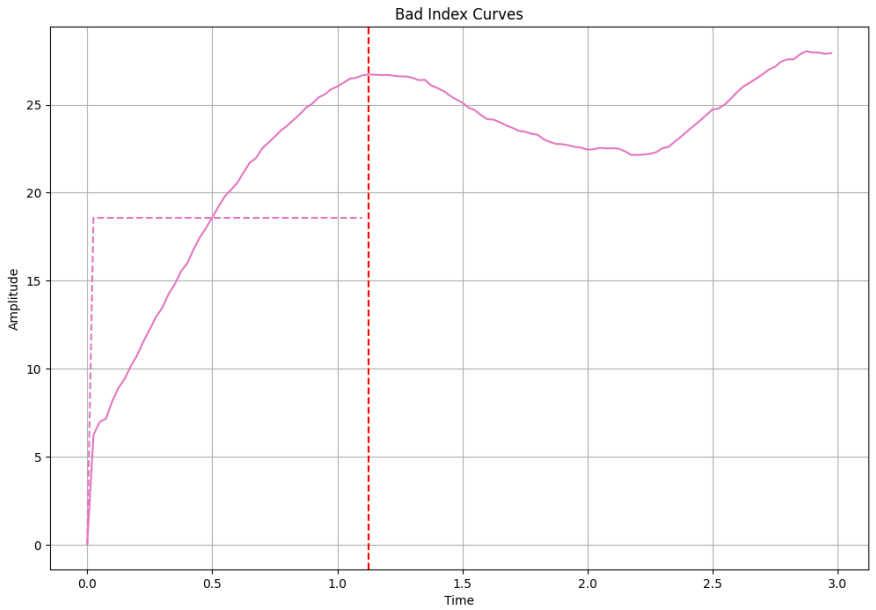
- I am fitting the early sections (up to the first maxima) of the coughing and breathing delay plots to an exponential function to find a characteristic timescale.
- Currently when the number of samples to the first peak is less than 5 samples, I set the value to zero.
- In some samples, frequency bins are spiking early causing these fixed zero values for timescale.

Exploration of Timescales of decorrelation onset:



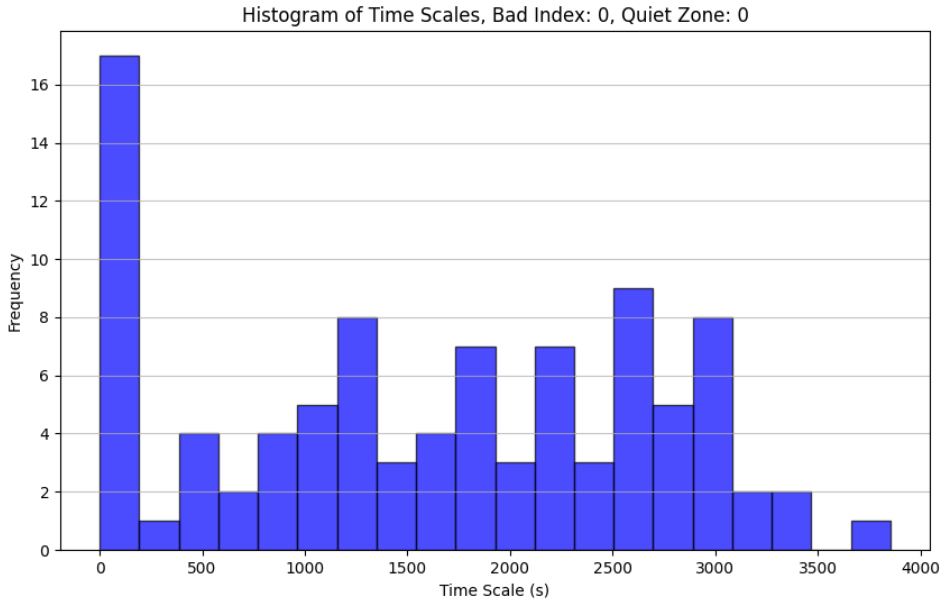
- These are some of the ‘bad plots’ causing the zero values.
- Clearly the maxima finding function is locating the early spike and missing the actual first ‘major’ local maxima.
- This could also be happening in scenarios even when the timescales aren’t getting flagged as bad indexes.

Exploration of Timescales of decorrelation onset:



- Upon investigating the instances of very low timescales I have noticed this trend.
- The fitting is clearly failing quite badly in a lot of instances.
- On the bright side, the first maximas are now very accurately being found with order parameter in argelextrema of 10 (0.25s either side).

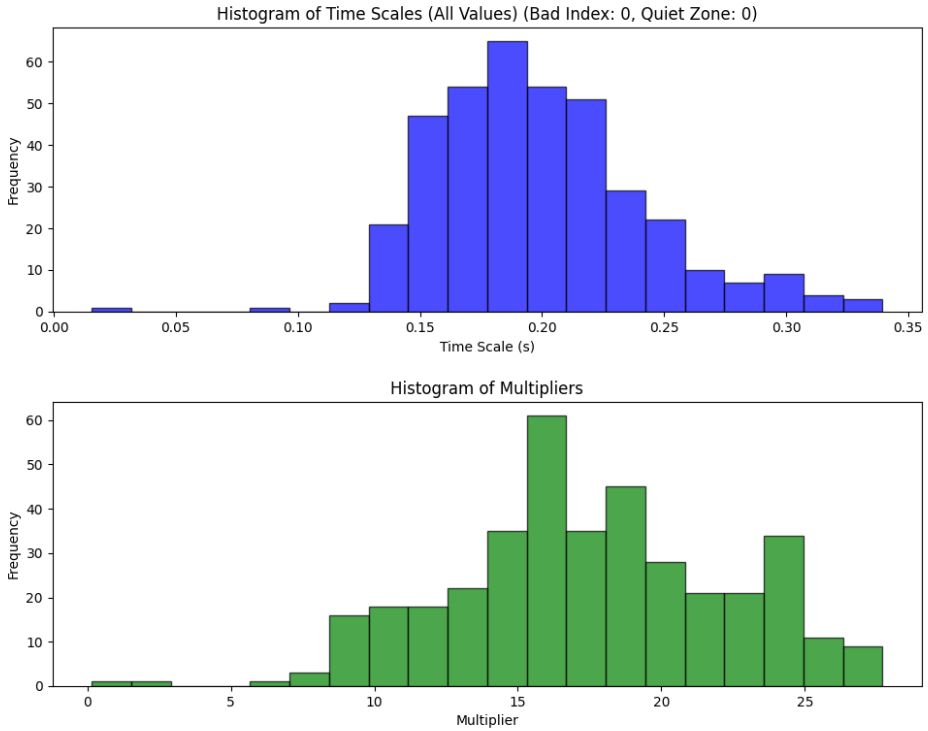
Exploration of Timescales of decorrelation onset:



- Tried using the stretched exponential fit
- f = a * (1 – exp(-x / b)) ** c
- Now the time scales are extending out much longer as in this case.
- Seems that the extra parameter is important in the fitting.
- I have also taken the delay spectrograms back to being normalised between 1 and 0. The error before was that the fitting couldn’t locate the optimal multiplier value a when initial guess was 1. This needs further exploration.

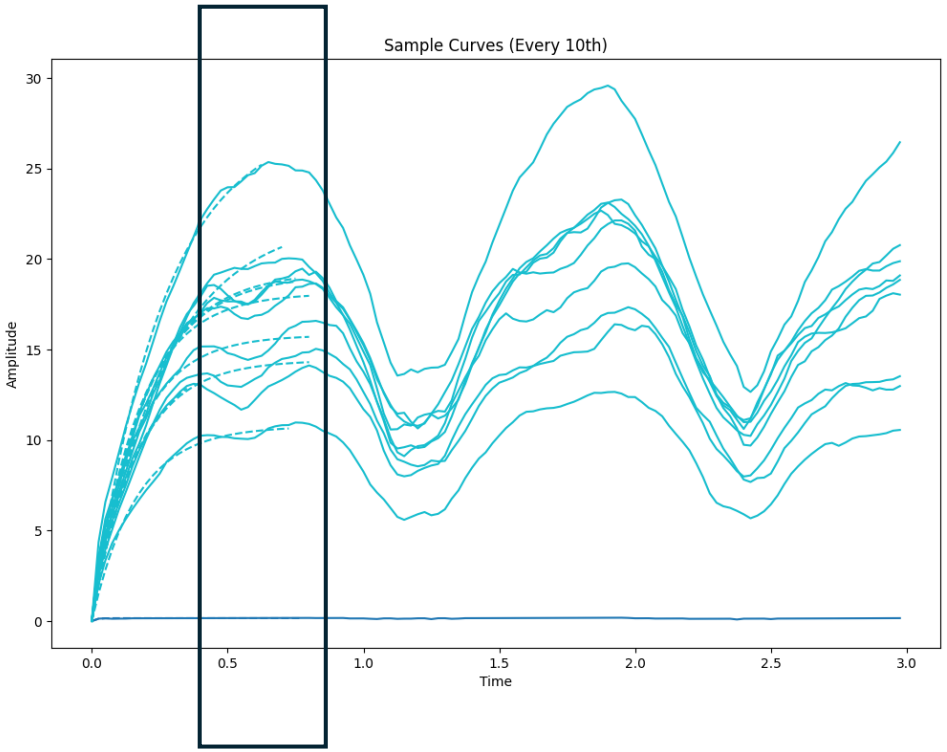
Modifying initial multiplier to 20 from 1 solves this.

Exploration of Timescales of decorrelation onset:



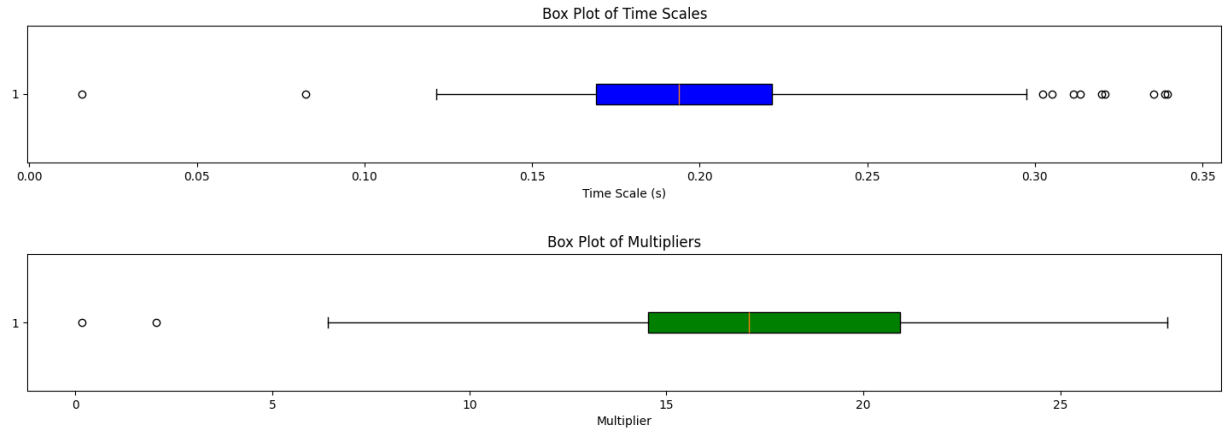
- Using just two parameters, the values are grouped much more nicely.
- I am going to try this way and use the mean median and standard deviation as input features.
- The overall fit of the exponential to the function isn’t as strong as the stretched exponential so will use that after having tried this.

Exploration of Timescales of decorrelation onset:



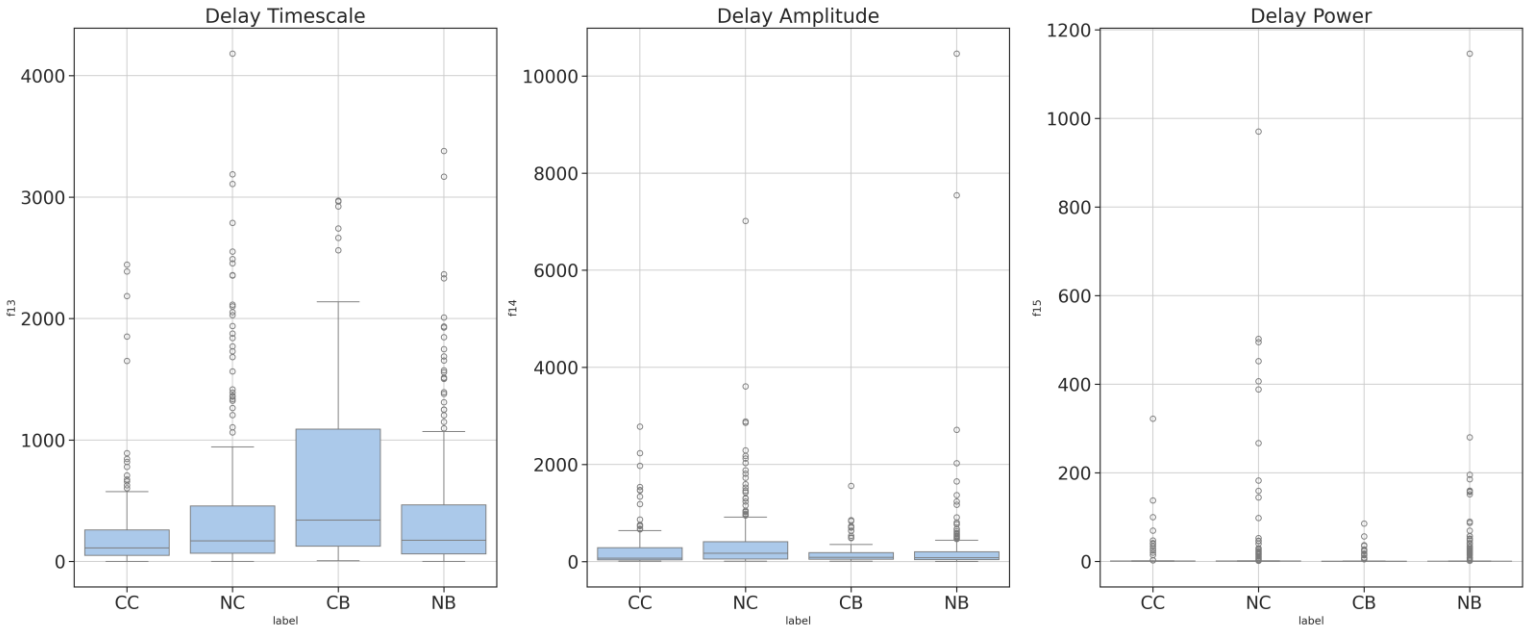
- Here are examples of the delay profiles at a range of frequencies between 0-3800Hz
- As shown, the double peak of the onset can be confusing to the local maxima locating function.
- Might need to reduce the order parameter further.

Exploration of Timescales of decorrelation onset:



- Box plot examples for the same delay plot as the last two slides.

Exploration of Timescales of decorrelation onset:

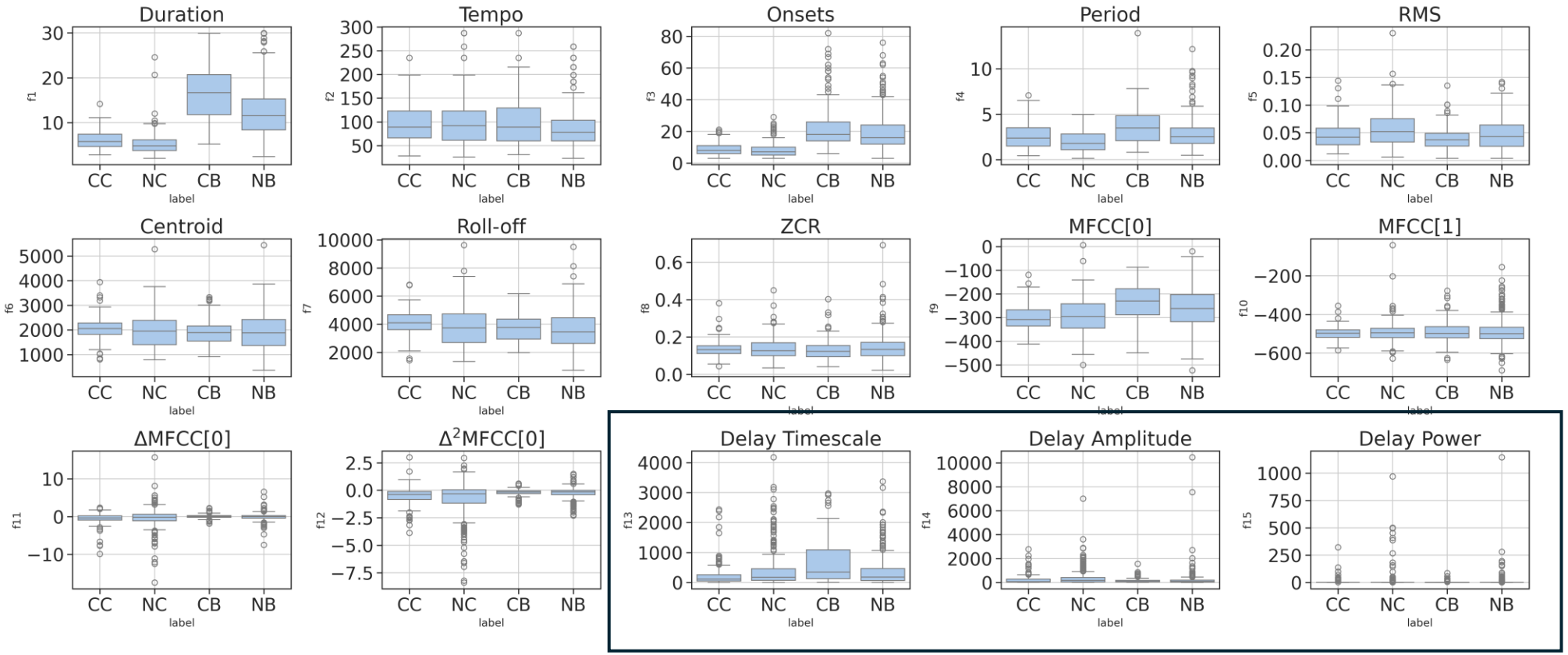


- Box plots of all the extracted features, using the three-parameter stretched exponential fit. The two parameter values were removing value from the classification.

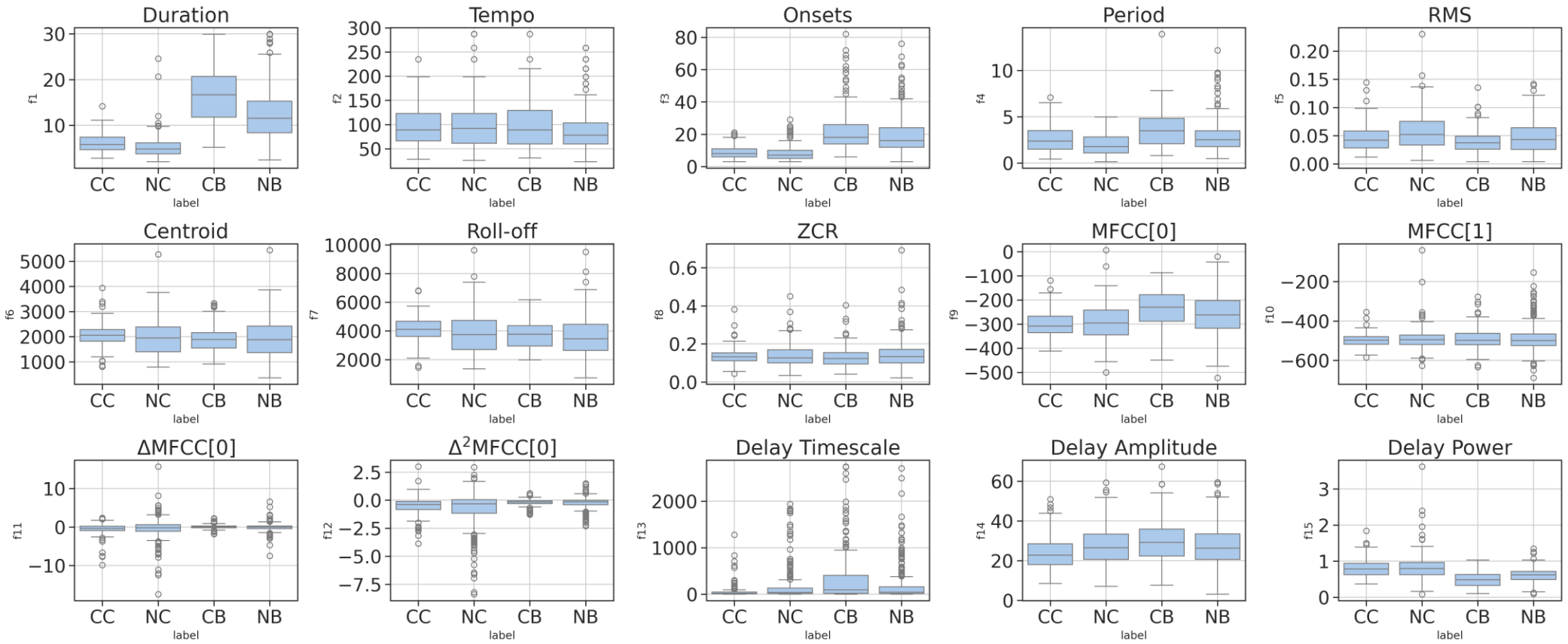
- I have also removed the frequency limit as I am just using statistical features and so can just ignore nonsensible bins.

- May next look to limit the features below certain thresholds as the wildly high values especially in amplitude will be hiding the useful time frames.

Full Feature Visualisation:

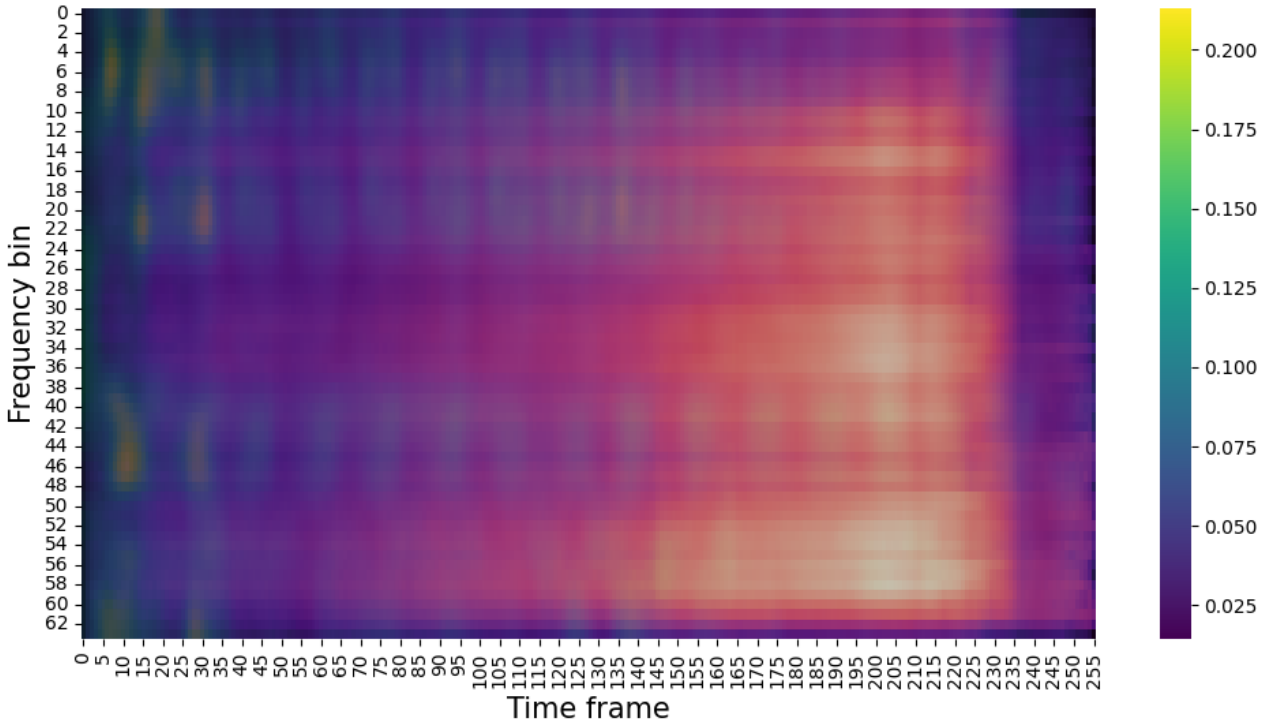


Full Feature Visualisation (removed large delay values):



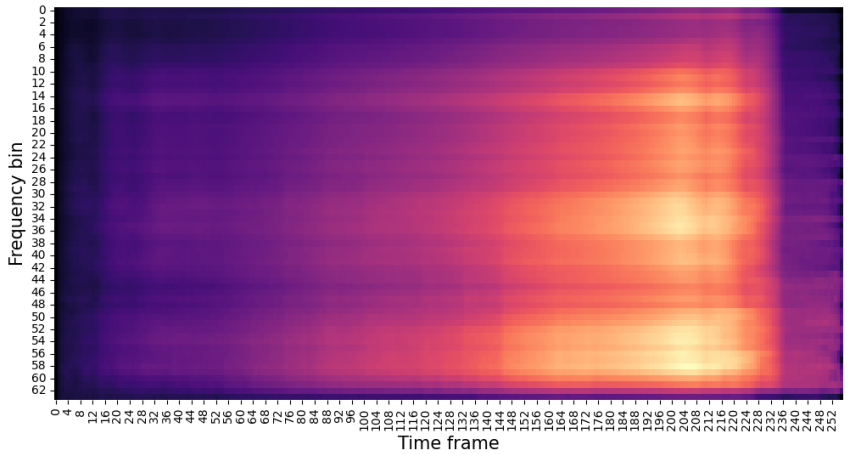
- Max Delay Timescale: 5000s
- Max Delay Amplitude: 100
- Max DelayPower: 10

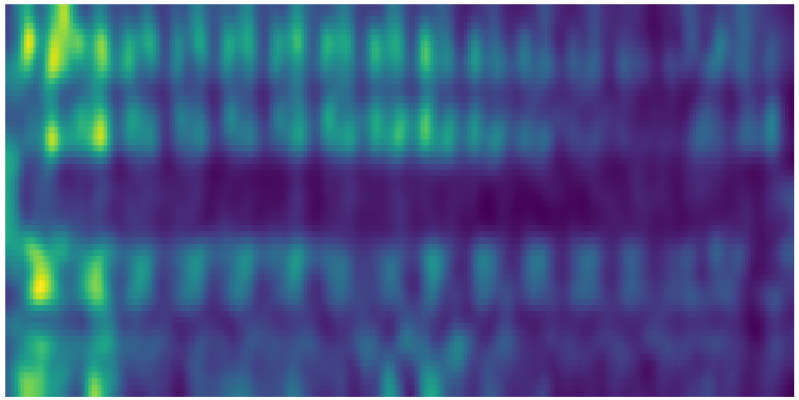
OPERA Saliency Maps:



- OwenGT, CovidUK
- Clear that model is paying attention to the earliest parts of the image as we expect.

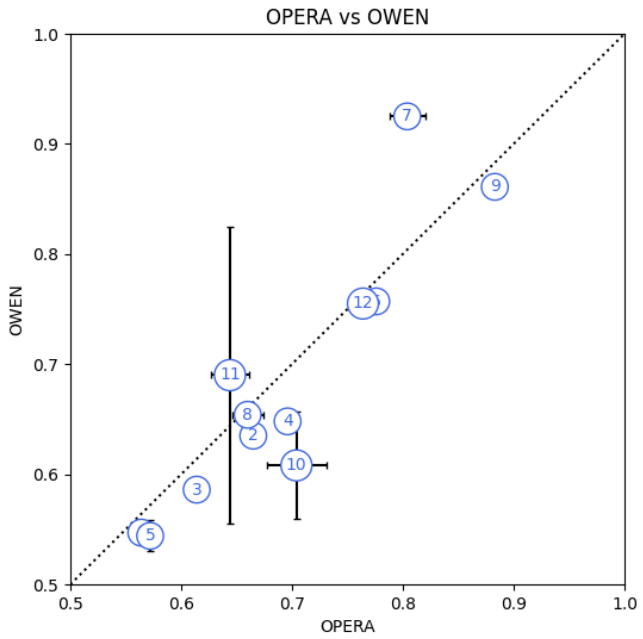
OPERA Saliency Maps:

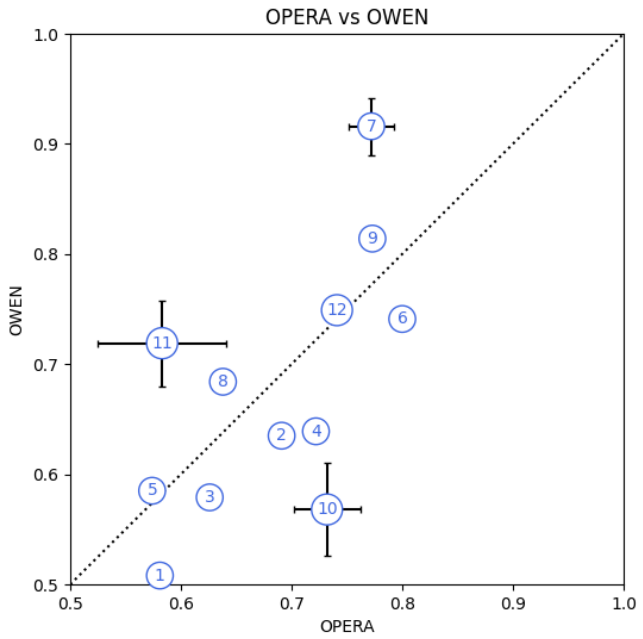


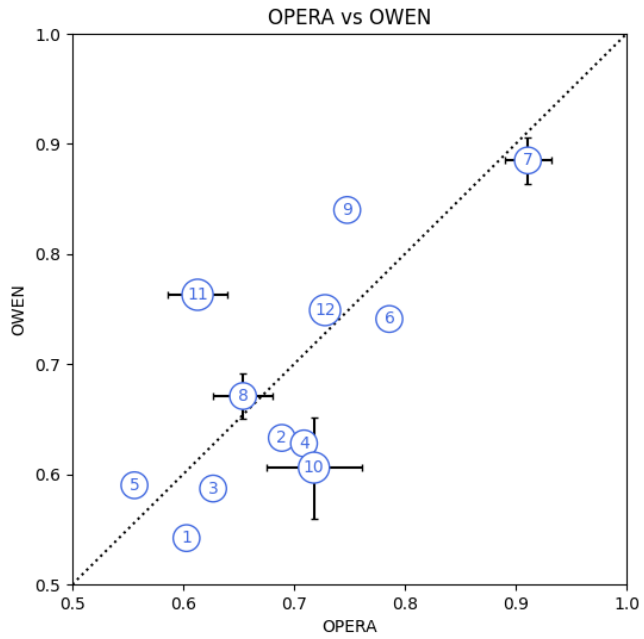


- OwenGT, CovidUK
- Saliency Map and Delay spectrogram pictured separately for more clarity.

OPERA vs OWEN plots:

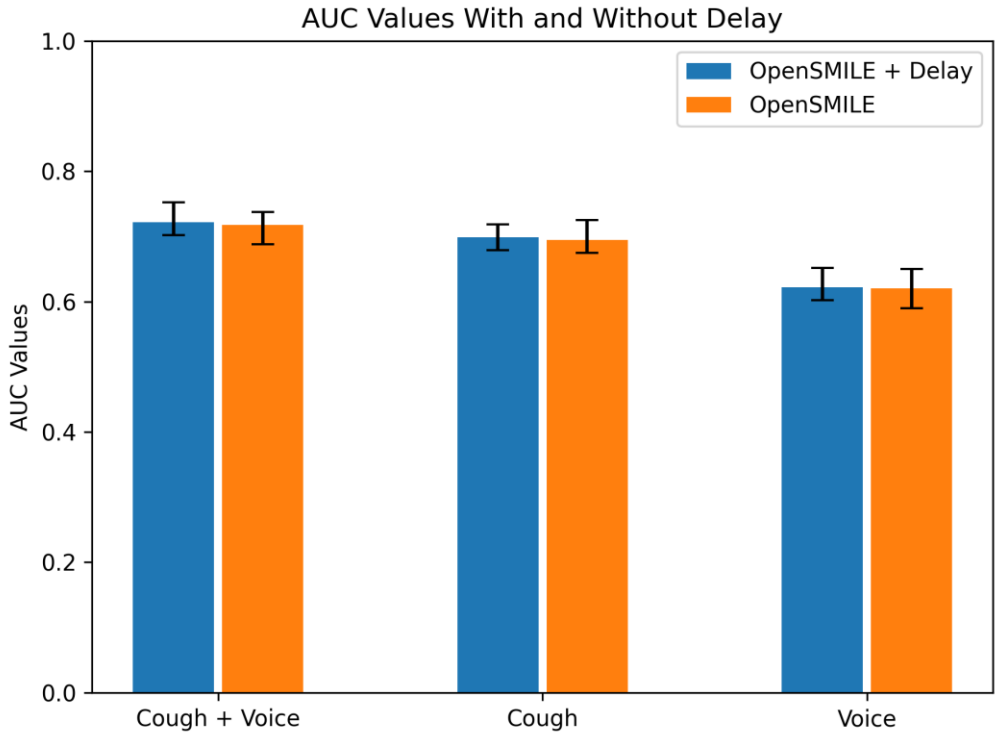


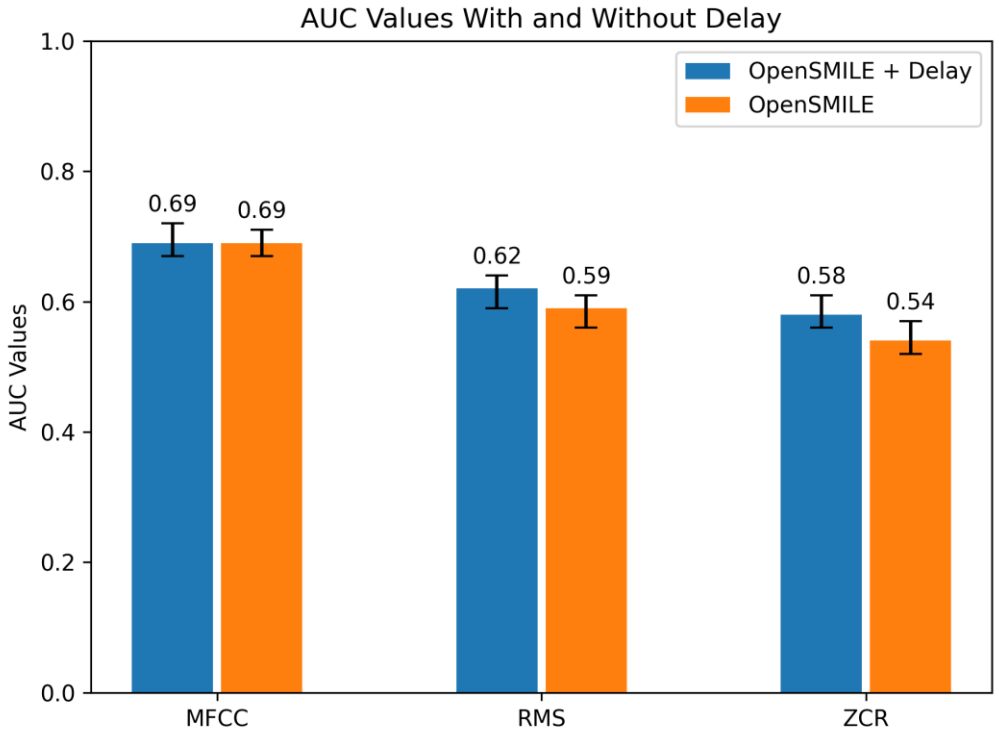




- Tasks 1-12 plotted with the training input time of 8s, 2s, 1s left to right.

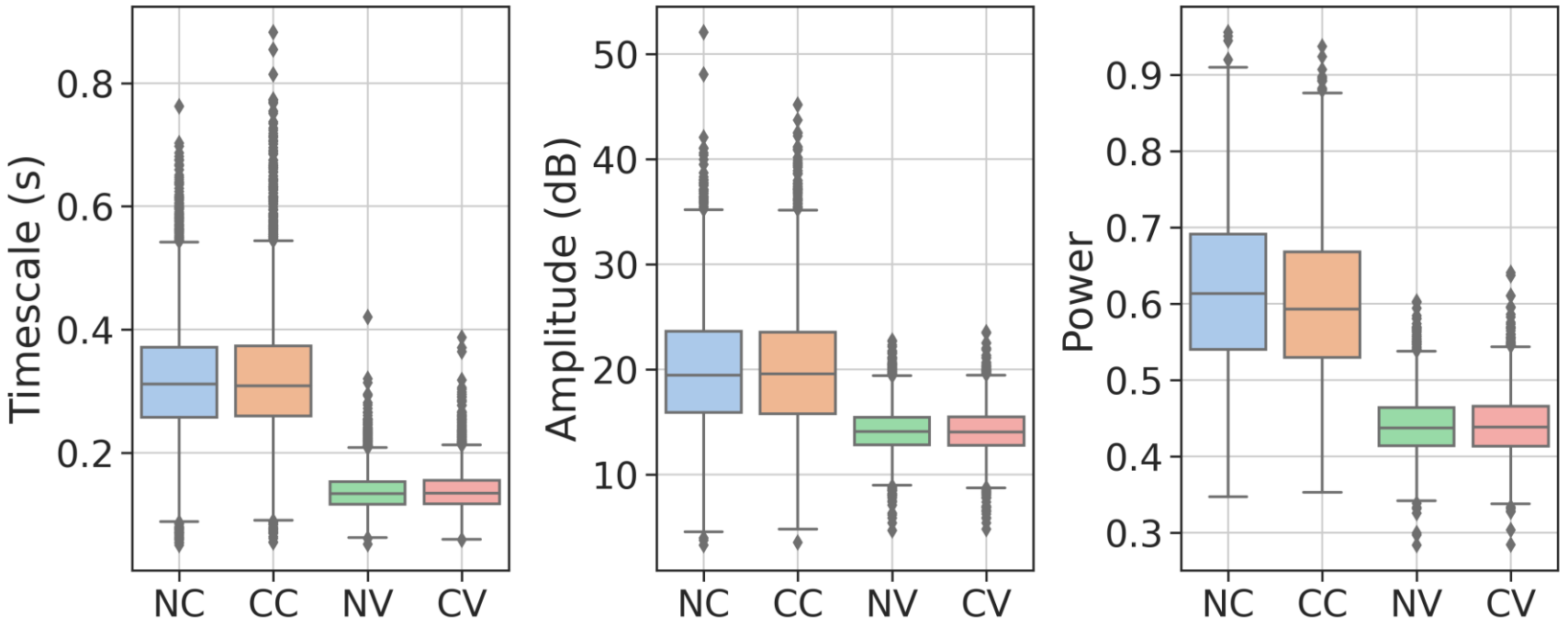
OpenSMILE with and without Delay Feature plots:



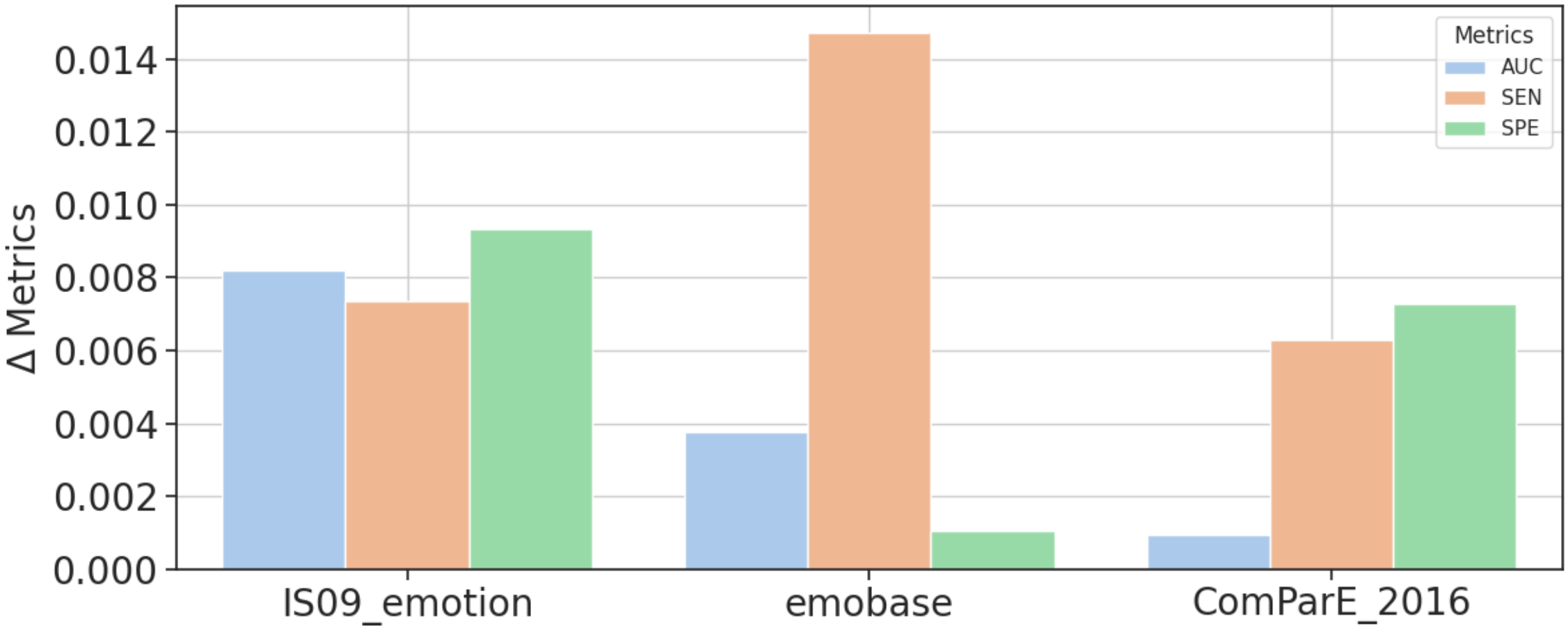


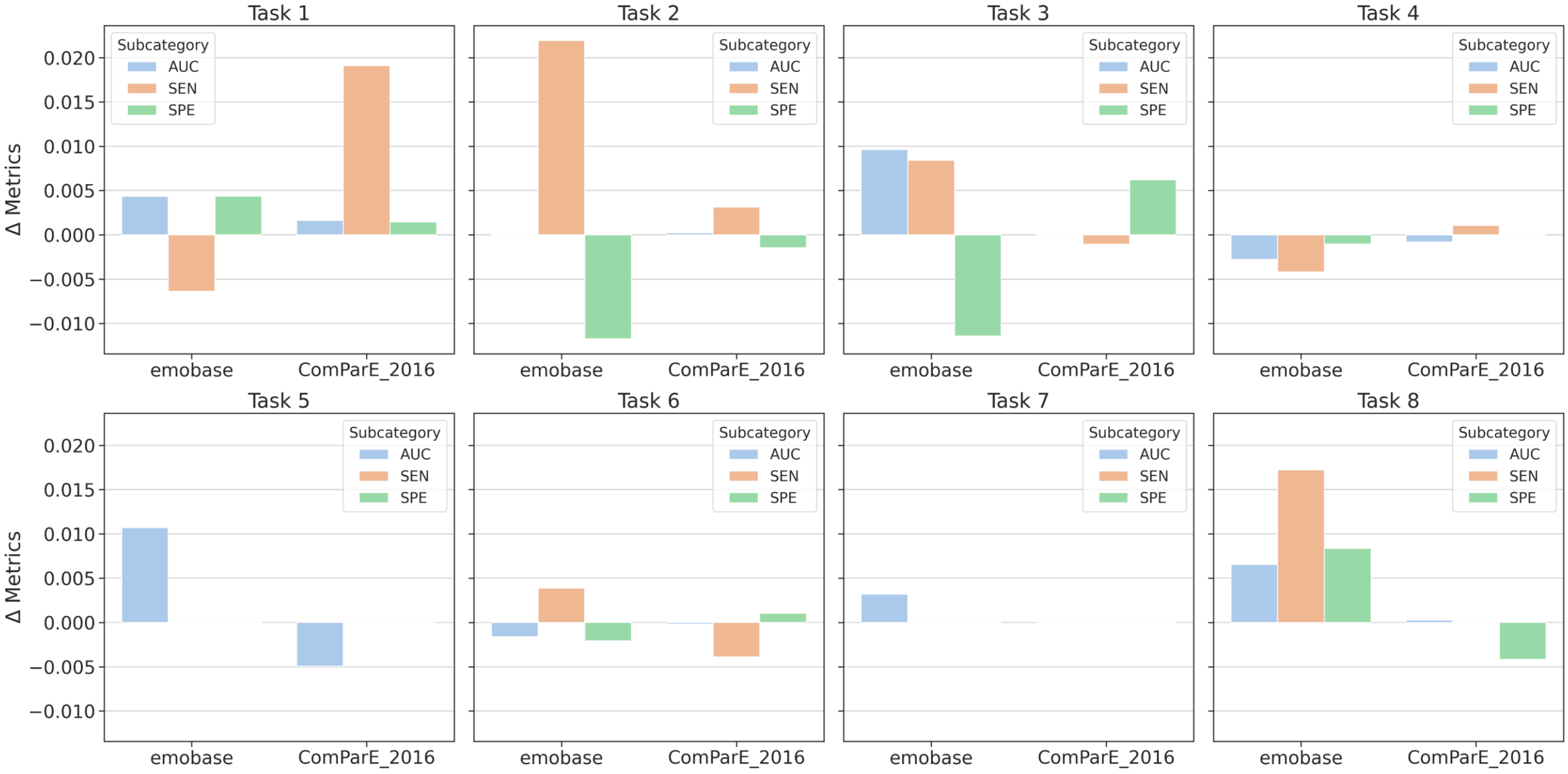
- Error bars showing the range between the 2.5% and 97.5% percentiles.
- Can see minor improvement with delay feature.
- The second plot only uses bot cough and voice data combined. There are more two more features used in the IS09_emotion feature pack but these were picking only the majority class so I have ignored.

Delay Feature Box Plots Covid19Sounds:

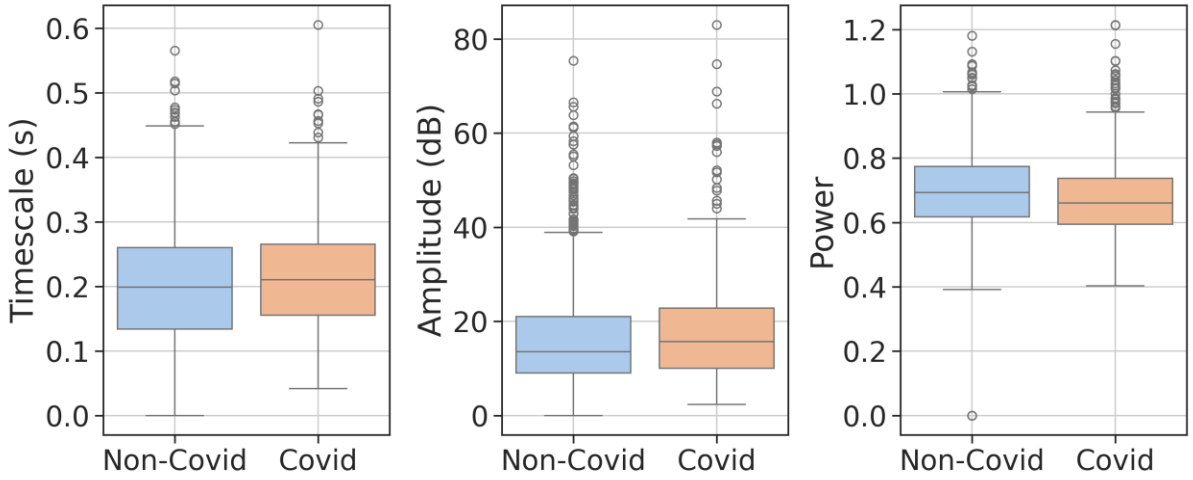


- Best separation comes from cough data with the power feature, this aligns with the results.

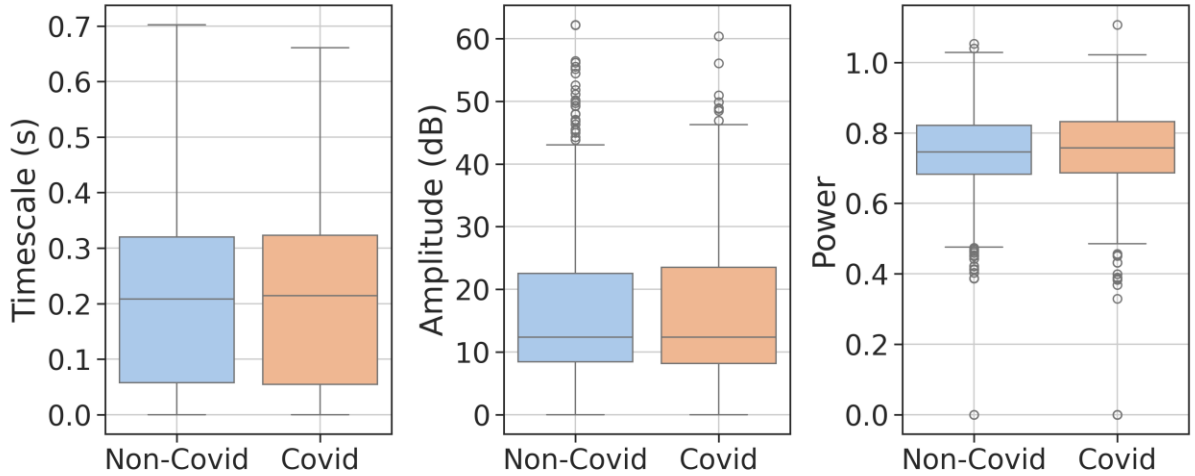

Metrics
AUC
SEN
SPE
Δ Metrics
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
IS09_emotion
emobase
ComParE_2016


Task 1
Task 2
Task 3
Task 4
Task 5
Task 6
Task 7
Task 8
Subcategory
AUC
SEN
SPE
Δ Metrics
0.020
0.015
0.010
0.005
0.000
−0.005
−0.010
emobase
ComParE_2016

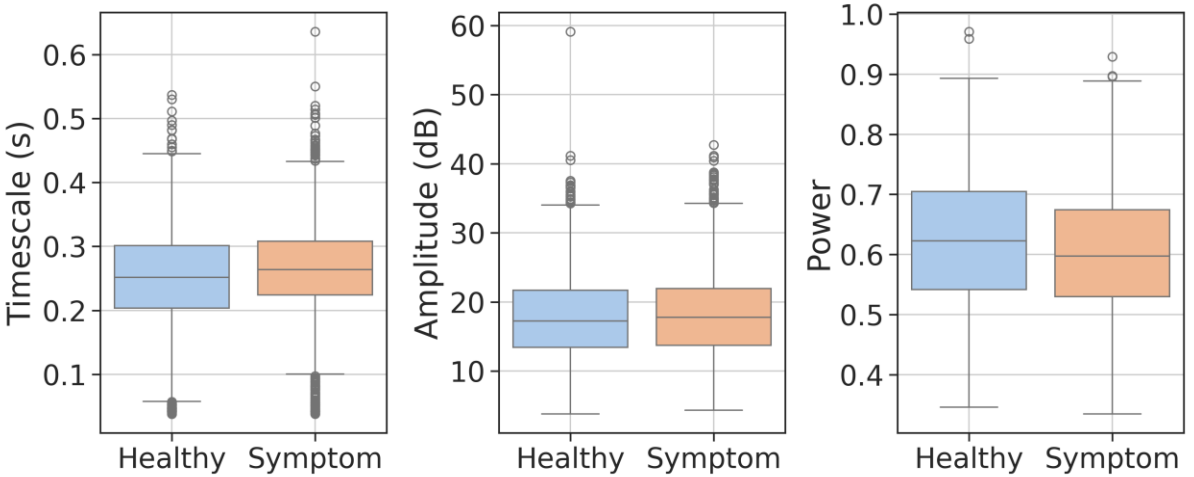
- Covid UK Cough



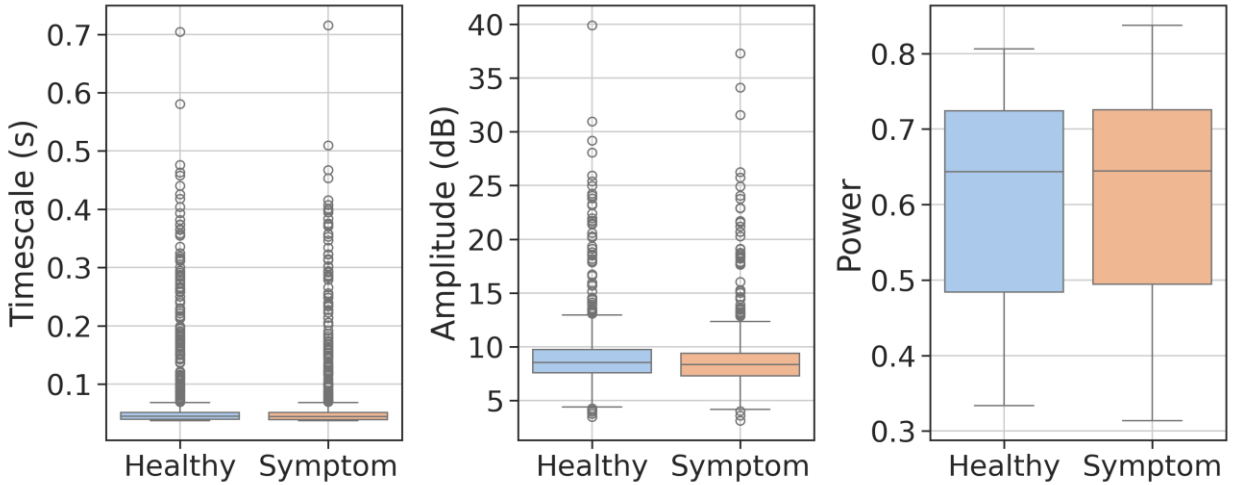
- Covid UK Breath



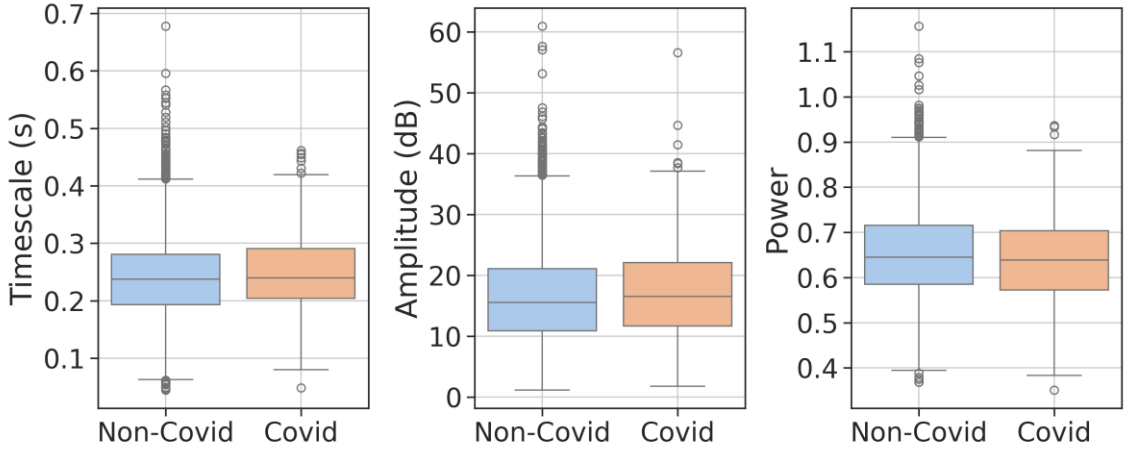
- Covid 19 Sounds Cough



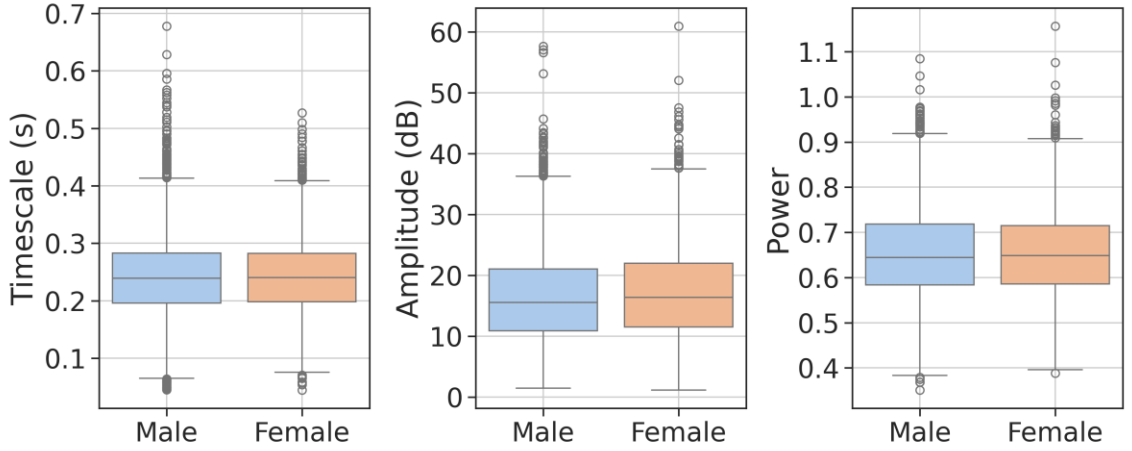
- Covid 19 Sounds Breath



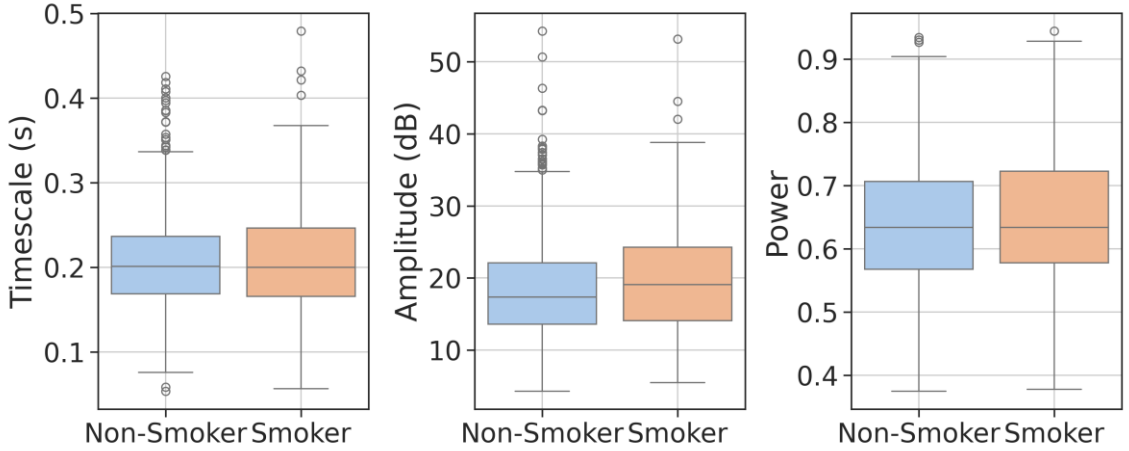
- Coughvid Cough Covid



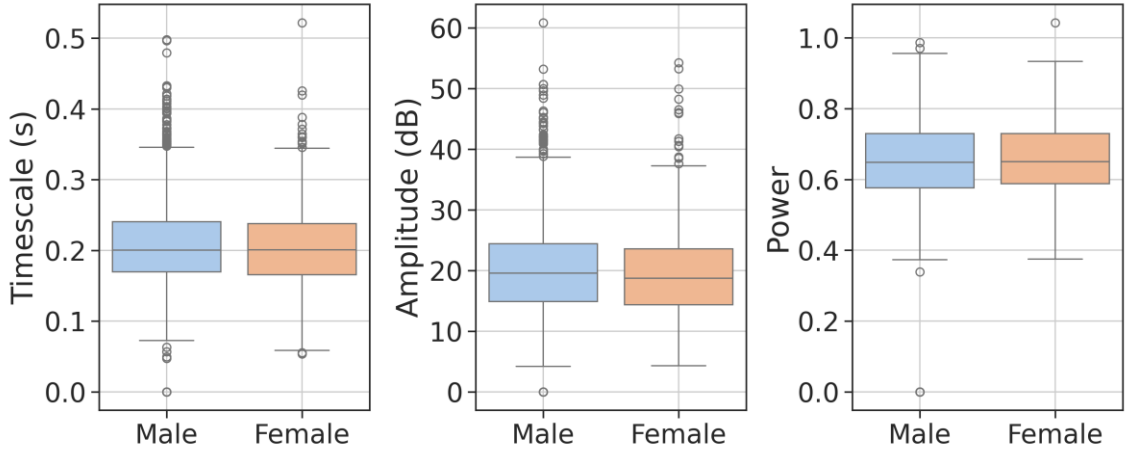
- Coughvid Cough gender



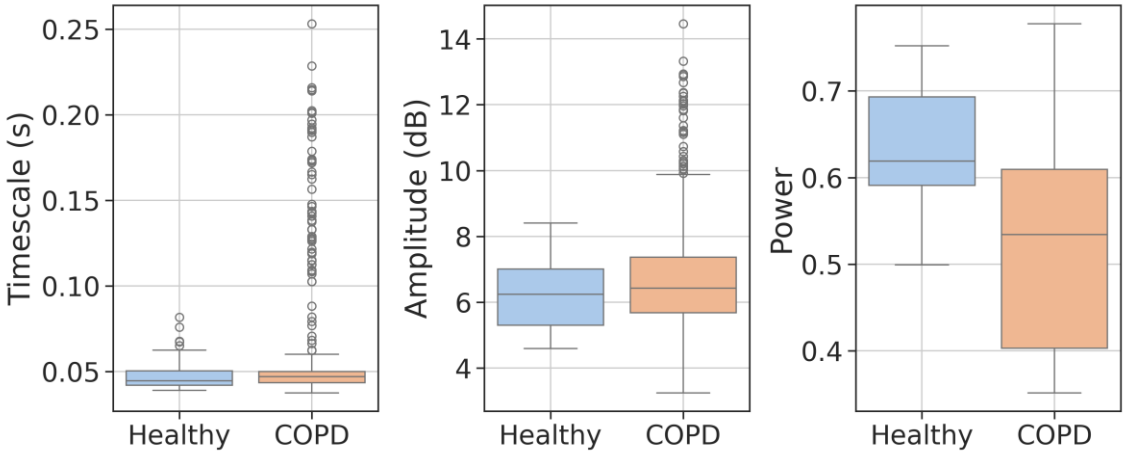
- Coswara Cough Smoker



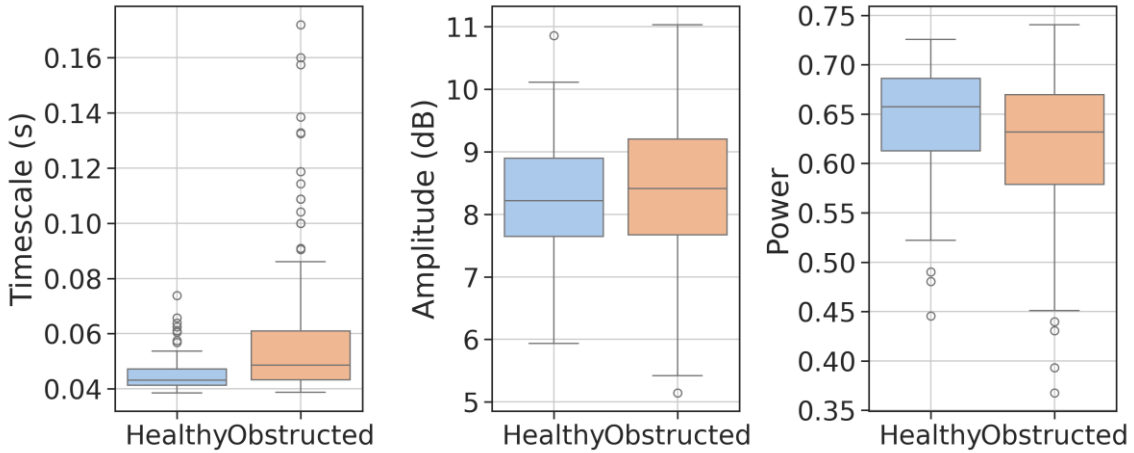
- Coswara Cough Sex



- ICBHI lung sounds COPD



- KAUH lung sounds asthma + copd



- Coswara Cough Smoker

- Coswara Cough Sex