

Prediction of a plucked bass: on the left can be seen how the different attack time values influence the predictions. The compression amount against the error for two cases is shown on the right.

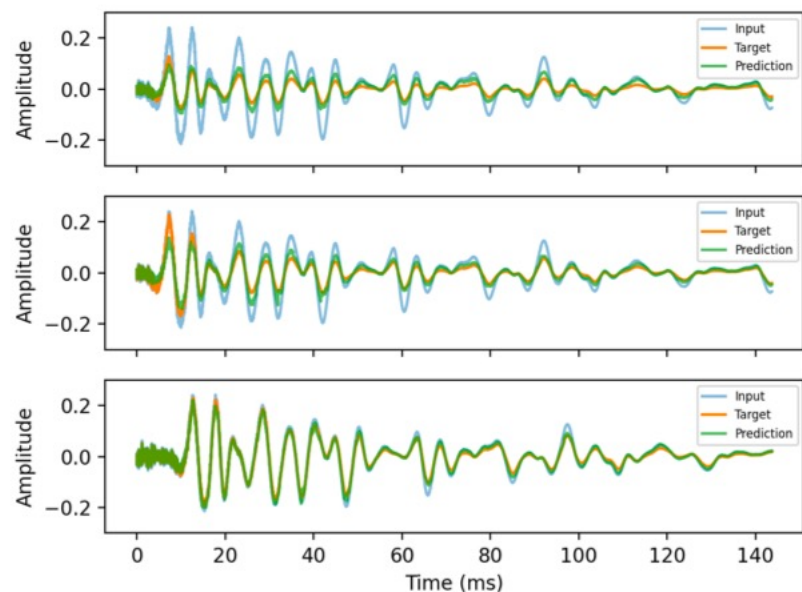


Figure 3: *Representative example of predicted waveform against the input target output for increasing values of attack and release time. Starting from the upper row: 0.5, 150, and 300 ms, for the attack time and 0.05, 5 and 10 seconds for the release time. The ratio and threshold are set to 6:1 and -20 dBU, respectively.*

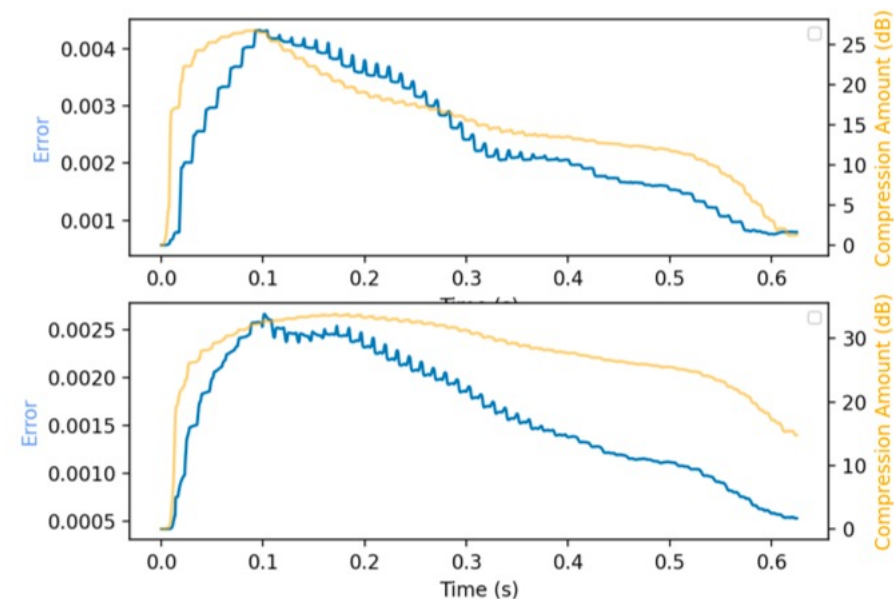


Figure 4: *Compression amount (i.e. gain reduction) against the error (RMSE) in two scenarios: 0.5 ms (upper) and 300 ms (bottom) as attack time. The images refer to two representative examples in which the model was fed with a plucked bass sound. The release, ratio, and threshold are set to 7.5 s, 6:1 and -20 dBU, respectively. The RMSE values are computed on a frame length of 4096.*

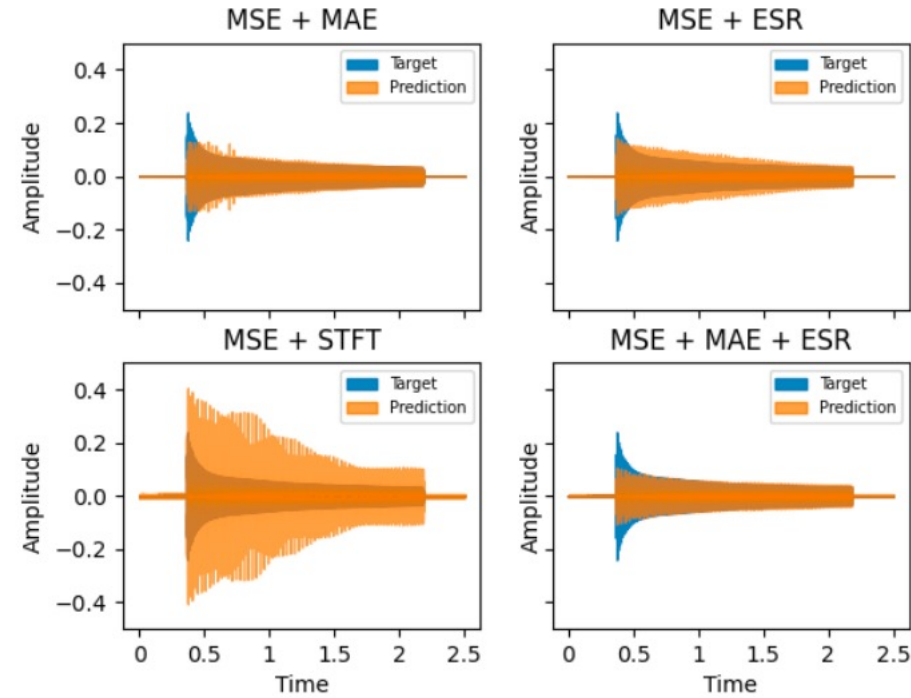
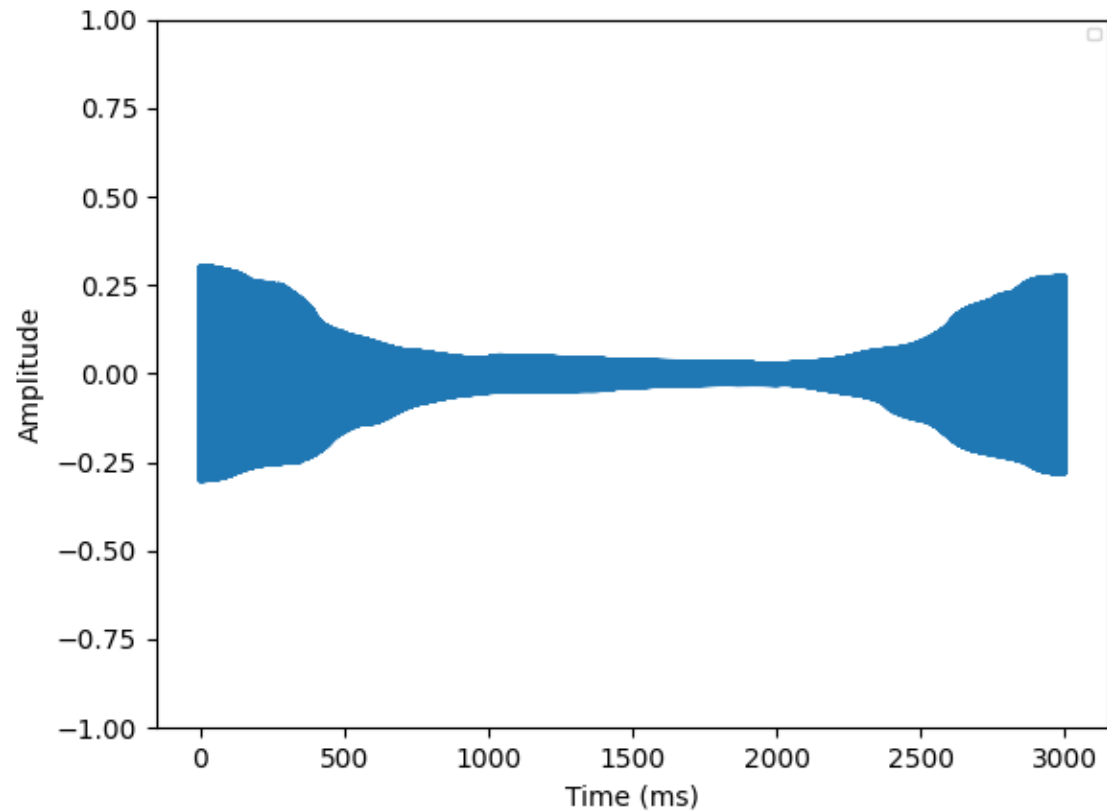


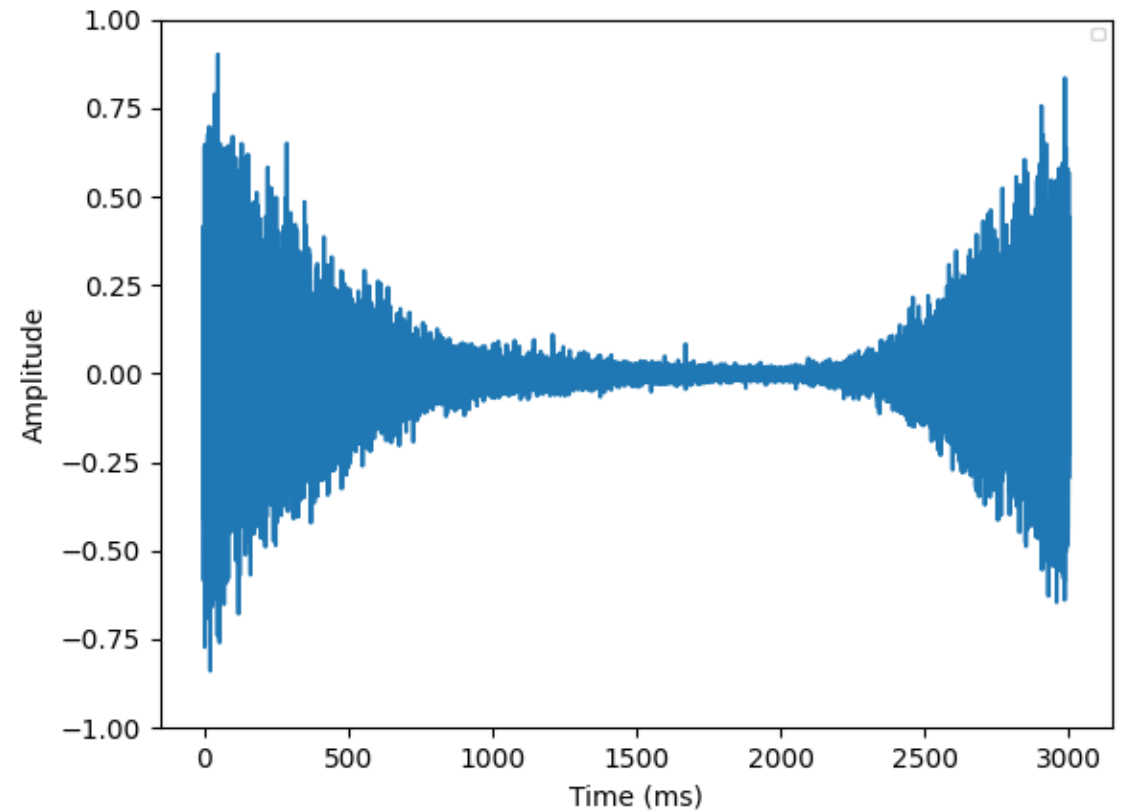
Figure 6: *Predicted waveform against the target for models trained using three combinations of loss functions: MSE + MAE, MSE + ESR, MSE + STFT-based. The target example refer to heavy-compressed scenario (-40 dBU as threshold and 10:1 as ratio)*

These results are used to determine the influence of the loss functions on the predictions and are not representative of the accuracy of the final model.

1000Hz Sinusoid



White Noise



*Audio generates when continuously varying the threshold and ratio parameters (every 16 samples).
The threshold initially increases from 0 to -40dBu; afterward, the ratio increases from 2:1 to 10:1.
In the second half, the process is reversed: ratio decrease to 2:1 and then threshold decrease to 0dBu.*