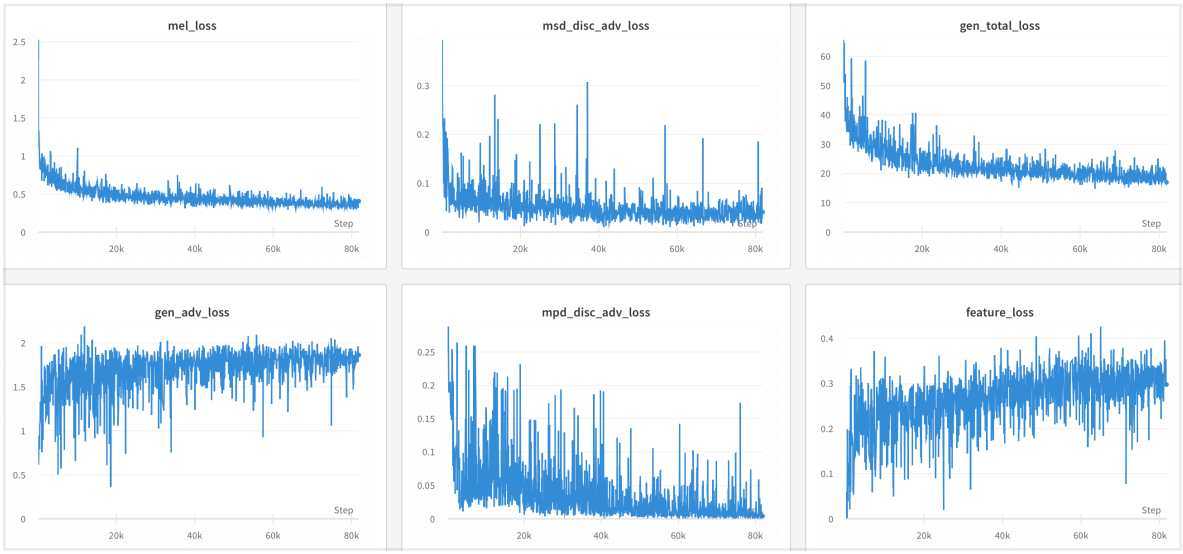
DLA
HW 4
NEURAL VOCODER BASED ON HiFiGAN

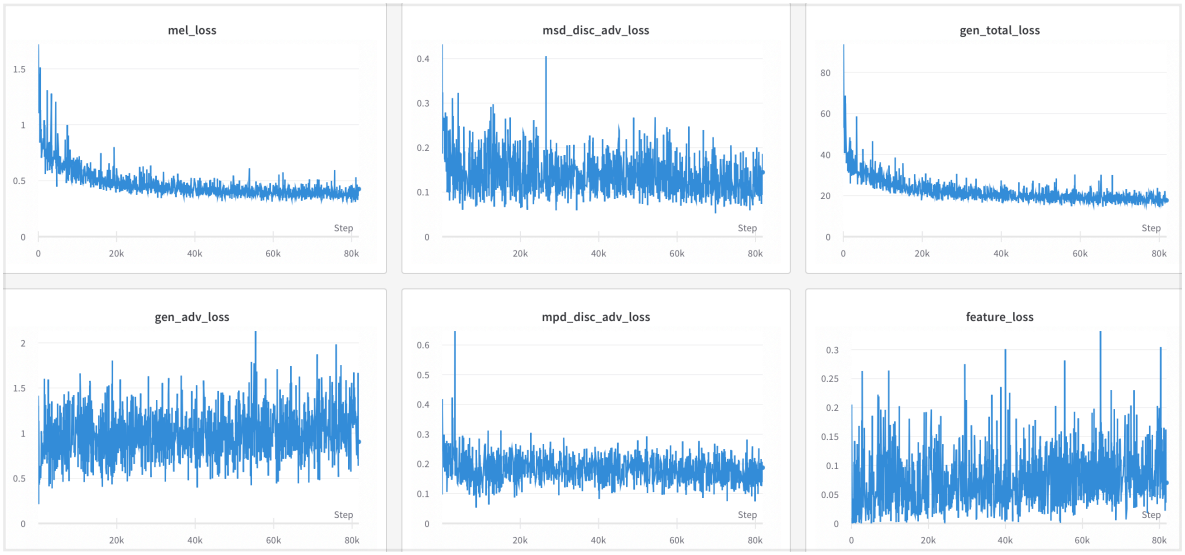
В домашней работе была реализована модель из статьи, которая обучалась на LJSpeech датасете.
Было проведено два эксперимента, первый из которых воспроизводил все, что было написано в статье, а второй - немного измененный вариант оригинального эксперимента из статьи. Оптимизаторы моделей AdamW - Decoupled Weight Decay Regularization
Генератор - Generator v3, как архитектура генератора, показавшая лучшие результаты из всех вариантов в статье.

Эксперимент 1
В первом эксперименте «выиграл» дискриминатор, это плохо, лучше, конечно, чтобы было не так. Согласно статье, идеальное решение достигается в случае, когда генератор генерит такие сэмлы, что дискриминатор может различить их от настоящих с вероятностью 0.5. В получившемся же первом эксперименте выиграл дискриминатор , соответственно, к таком сойтись невозможно, и весь adversarial подход становится бессмысленным.
Посмотрим на логи:



Личные впечатления: текст понимаю, интонация похожа на оригинальную аудиодорожку, шум есть, несильно мешает различать речь.

Эксперимент 2
Поэтому во втором эксперименте обновляли веса дискриминатора на n-ную итерацию только (n=5).
Дискриминатор не выиграл, и оба и генератор, и дискриминаторы показывают примерно константный лосс с какого-то момента, что часто и возникает при правильном обучении ганов.
Параметры те же самые.
Посмотрим на логи.



Личные впечатления: текст понимаю, интонация также похожа на оригинальную аудиодорожку каждого тестового файла, однако присутствует шум, который ассоциируется с «голосом робота»)
Шум, как и в результатах первого эксперимента, не мешает различать то, что сказано, хоть и «более правильное» не убрало этот шум.

Все-таки в первом эксперименте меньше шума и мне больше нравятся аудиодорожки: речь там четче, «приятнее» уху.