

## Данные

Перед началом обучения чего-либо, получил мел-спектрограммы из транскрипций аудиофайлов с помощью TextToSpecConverter.

## Модель

Взял MelGAN модель и реализовал ее в pytorch\_lightning. Но уменьшил кол-во параметров, потому что не хватало ресурсов, по этой же причине очень маленькие батчи были. Добавил lr\_scheduling еще. Немножко игрался с весом для feature\_loss. Сначала поставил 10, потом 8, с 8 было лучше.

## Результаты

Поначалу получался просто какой-то шум, но с увеличением кол-ва эпох начало получаться что-то похожее на речь, отдельные словосочетания можно распознать, даже если не знать какой текст пытаемся сгенерировать. Я обучал 30 эпох, ~13 часов. Наверное, если взять модель побольше и пообучать дольше, то могут получиться результаты лучше.