# Отчет по TTS

## ДА, Я ИСПОЛЬЗОВАЛ НОРМАЛЬНЫЕ АЛАЙНМЕНТЫ

Общий Wandb - https://wandb.ai/pos/tts_project?workspace=user-pos

## Интересные эксперименты:

- Со старым алайнером, батч 64. Обучалось 16.5 тысяч итераций, оптимайзер `Adam(lr=4e-3, betas=(.9, .98), eps=1e-9)` (почти как в статье Attention Is All You Need), scheduler взят по описанию из этой же статьи. Wandb по эксперименту (тогда еще не логировал каждый лосс по-отдельности). Результаты были не очень хорошие, некоторые слова не поймешь без подписи, решил попробовать нормальный алайнер.

- С алайнментами из https://github.com/xcmyz/FastSpeech, батч 4. Обучалось сначала 20 эпох (по 3275 итераций) с тем же шедулером, lr = 5e-2. Wandb. Веса промежуточного результата в файле fastspeech. По графику лосса заметил, что если продолжить обучать, начав с последнего значения lr (2e-4), то можно попробовать уменьшить лосс на длительностях. Продолжил, scheduler: `StepLR(step_size=3275, gamma=0.8)`, обучал еще 10 эпох - в записях стало немного меньше шума, голос стал громче, лоссы изменились не сильно. График - Wandb. Веса в файле fastspeech_new.

- Другой способ уменьшить лосс на длительностях - увеличить kernel_size в DurationPredictor'е (5 и 7 вместо 3 и 3) и уменьшить в нем dropout (0.1 -> 0.05), batch_size = 16. График - Wandb. Лосс, конечно, убывает быстрее (принимая во внимание то, что батч стал в 4 раза больше), однако звук более зашумленный и качество ощутимо ниже, чем во втором интересном эксперименте. Лосс на спектрограммах, кстати говоря, даже выше, и более нестабильный. Веса в файле fastspeech_16.

# Результаты

Лучшим по качеству, с субъективной точки зрения, стал продолженный второй эксперимент. В нем есть проблема с произношением слова дефибриллятор (всего пару раз за все время обучения модель смогла нормально расставить интонации), однако несмотря на это общее качество довольно хорошее, другие слова четкие и громкие, хоть и присутствуют некоторые синтетические резкие звуки.