

* Explain some worst bugs that you had to fix.

First, train dataloader works only for the first epoch. Then there is no data. It seems to happen because of generator that is used in batch sampler.

Second, alignment model shapes weren't obvious sometimes.

* Tell your story of training runs. ****Attach screenshots of loss curves.**** How many runs in total have you done? What was the longest one? Did you have to use optimization tricks/heuristics/magic from the Web?

I haven't done any training run. Just checked that model works for one batch.

* Was it easy to obtain the required audio quality? Do you think the best model could improve if trained for longer?

Don't know.

* For you personally, what were the ****top 3 hardest things**** in this assignment? (name anything, even "*waiting to be unblocked on Colab*")

1. Dataloader troubles.
2. Troubles with downloading datasets (it took a lot of time, so I've just downloaded pickle files and used it later).
3. Alignment model meaning.

* What have you learned while doing this task?

Using transformers, the pipeline of text2speech modelling.

* What questions of yours has this task provoked? E.g. "I was curious if Transformers are used in X" or "I'm wondering if real papers really don't mention Y".

* How would you improve this assignment?

Make at least one training run :)